

Robust Portfolio Choice*

Valentina Raponi Raman Uppal Paolo Zaffaroni

9 November 2018

Abstract

We investigate the effect of model misspecification on mean-variance portfolios when investors have a preference for robustness, in the sense of Hansen and Sargent (2008). Our analysis is founded on the Arbitrage Pricing Theory (APT), which we exploit along three dimensions. First, we show how the APT can be extended so that it captures models not just with small pricing errors unrelated to factors but also *models with large pricing errors* from mismeasured and missing factors. Second, we show that the APT restriction is critical for identifying the *set of possibly misspecified models* considered by the investor. Third, the APT allows one to construct the optimal portfolio *robust to model misspecification* arising from the incorrect specification of the distribution of factor risk premia, leading to new insights about mean-variance portfolios when the number of assets is large. We illustrate how our theoretical insights lead to a significant improvement in the out-of-sample performance of mean-variance portfolios using simulations.

JEL classification: G11, G12, C58, C53.

Keywords: Model misspecification, actor models, factor investing, alpha, smart beta.

*Raponi is affiliated with Imperial College Business School; email: v.raponi13@imperial.ac.uk. Uppal is affiliated with EDHEC Business School and CEPR; email: Raman.Uppal@edhec.edu. Zaffaroni is affiliated with Imperial College Business School; email: P.Zaffaroni@imperial.ac.uk. We thank the Mathematics Department at Imperial College London for allowing us to use their computer cluster. We gratefully acknowledge comments from Dante Amengual, Turan Bali, Harjoat Bhamra, Tim Bollerslev, Svetlana Bryzgalova, Timothy Christensen, John Cochrane, Rama Cont, Victor DeMiguel, Serge Darolles, Patrick Gagliardini, René Garcia, Lorenzo Garlappi, Christian Gourieéroux, Amit Goyal, Christian Julliard, Andrew Karolyi, Ralph Koijen, Michelle Lee, Xioaji Lin, Lionel Martellini, Vasant Naik, Stefan Nagel, Ľuboš Pástor, Markus Pelger, Hashem Pesaran, Eric Renault, Cesare Robotti, Lucio Sarno, Olivier Scaillet, Jay Shanken, Yazid Sharaiha, Eric Shirbini, Claudio Tebaldi, Nikolaos Tassaromatis, Allan Timmermann, Fabio Trojani, Tan Wang, Bas Werker, Motohiro Yogo, and Guofu Zhou.

1 Introduction

The mean-variance model of portfolio choice is the cornerstone of financial economics because it captures the fundamental tradeoff between expected return and risk.¹ To implement this model a decision maker needs to know the expected returns, variances, and covariances of the assets. However, when making financial decisions, it is not possible for the decision maker to wait for an infinite amount of data to resolve her fear of model misspecification by estimating these moments of asset returns nonparametrically, i.e. by estimating the sample means and covariances, or to estimate directly the portfolio weights, as in Brandt (2004); moreover, while one can estimate covariances with reasonable precision by increasing the sampling frequency, this is not possible for expected returns (Merton, 1980). This motivates the use of a fully parametric model, such as the Capital Asset Pricing Model. However, as highlighted by Hansen and Sargent (1999), “*with tractability comes misspecification.*” For instance, in our context, when specifying a particular factor model for asset returns, one or more priced factors could be mismeasured (Roll, 1977) or missing (Pástor, 2000).

In this paper, we investigate the effect of model misspecification on mean-variance portfolios when investors have a preference for robustness. We consider a two-stage approach for identifying a portfolio rule that is robust to misspecification. In the first stage, the decision maker (investor) constructs an *approximating* model. The first stage is crucial because the robust portfolio will then lie in a small neighborhood around the approximating model, and therefore, the performance of the robust portfolio will depend on the choice of the approximating model. In contrast to the robust-control literature, which is typically silent about the process by which an agent arrives at her approximating model, we use asset-pricing theory to identify the approximating model for asset returns. In the second stage, the investor recognizes that the model may be misspecified, and therefore, uses both robust control theory, along with our new insights when the number of assets is large, to choose a portfolio that is robust to various forms of misspecification described below.

The first stage of our analysis, where we construct the approximating model, is founded on the Arbitrage Pricing Theory (APT) of Ross (1976). However, in contrast to the existing literature, we use the APT not just as a *given* model of asset returns, but to define the *set of approximating asset pricing models*, possibly misspecified, large enough to contain the true

¹To see the broad range of applications of mean-variance theory, see Cochrane (2014).

model generating asset returns.² To encompass a larger set of models, our first contribution is to extend the APT so that it captures not just “small” pricing errors but also “large” pricing errors related to mismeasured or missing pervasive factors. This contrasts with the existing APT literature, where the pricing errors are small and *unrelated* to factors; see, for example, Cochrane (2005, Ch. 9.4).

The first stage of our analysis consists of estimating the approximating model, which lies in the set of models spanned by what we refer to as the *extended APT*. We illustrate the estimation using constrained-maximum likelihood (ML), where the constraint is the no-arbitrage restriction of the APT on the weighted sum of squared pricing errors. In particular, imposing the APT restriction when estimating the parameters ensures that the “approximating model” lies within the set of no-arbitrage models. Our estimation procedure takes into account various forms of misspecification potentially affecting the approximating model; for instance, missing factors, mismeasured factors, and pure pricing errors unrelated to factors. In the case in which the pricing errors contain *both* latent factors and a component unrelated to latent factors, the APT no-arbitrage restriction plays a second, even more fundamental, role: in the absence of this restriction, the model is not (econometrically) identified, and hence, cannot be estimated. In fact, we show that the role of the APT restriction goes beyond that of achieving identification in the presence of missing factors: it ensures identification also of models with *non-traded* observed factors and pure pricing errors. This part of our work extends the rich insights of MacKinlay and Pástor (2000), who study estimation of models with missing factors.

The second stage of our analysis consists of choosing a robust portfolio building on the max-min approach of Gilboa and Schmeidler (1989). Hansen and Sargent (2008) provide a comprehensive discussion of the notion of robust decision making by economic agents who regard their model as an approximation and who desire decision rules that work over a set of models in the neighborhood of the approximating model. In practice, the max-min approach determines robust portfolio weights by minimizing the mean-variance objective

²The APT is particularly well-suited for our purpose because it allows for the possibility of model misspecification, and hence, mispricing (alpha), while still imposing no arbitrage. Moreover, the APT is a very general asset-pricing model that can accommodate a variety of observed (traded or nontraded) factors. The factors could be statistical, for example, based on a principal-component decomposition of returns; macroeconomic, for example, shocks to inflation, interest rates, and exchange rates; or, characteristic-based, for instance, industry, country, size, value, return momentum, and liquidity. For further details of the variety of applications of the APT, see Connor, Goldberg, and Korajczyk (2010, Ch. 4–6). For the importance of factor investing, see the excellent discussion in Ang (2014).

function over the set of pricing errors subject to the constraint that these pricing errors are not statistically distinguishable (at conventional significance levels) from the estimated pricing errors associated with the approximating model. Then, the mean-variance utility is maximized over the choice of portfolio weights.

In particular, we show that, under the APT, the robust mean-variance portfolio can be decomposed into an “alpha” portfolio, which depends only on pricing errors with zero exposure to common risk, and a “beta” portfolio, which depends on factor risk premia.³ We treat misspecification in the alpha portfolio via the max-min approach described above. On the other hand, we treat misspecification in the beta portfolio using our new results: we show that, when the number of assets is large, the weights of the alpha portfolio typically dominate the weights of the beta portfolio.⁴ We then show that, under a set of mild conditions, the beta portfolio can be replaced, *without any loss of performance*, by a class of benchmark portfolios (such as the equal-weighted or value-weighted portfolios) that by construction are *functionally independent* of the mean vector (risk premia) and covariance matrix of the observed factors, and hence, completely immune to beta misspecification.

Finally, we demonstrate how these new insights can and should be used to improve the estimation of the return-generating model and the portfolio weights in the presence of model misspecification. Using simulations, we show that it is possible to take advantage of our theoretical insights to achieve economically and statistically significant improvement in the out-of-sample performance of mean-variance portfolios.

The rest of the paper is organized as follows. In Section 2, we discuss the literature related to our work. In Section 3, we specify the linear factor model for asset returns, summarize the results in the existing literature for the APT, extend the APT to the case of large pricing errors related to missing pervasive factors, and explain how to use the APT to construct the set of approximating models. In Section 4, we specify we describe the robust portfolio choice problem and explain how to mitigate model misspecification in

³This decomposition is important because the world’s largest hedge funds, such as Bridgewater Associates, offer alpha and beta portfolios. Similarly, sovereign-wealth funds, such as Norges Bank, separate the management of their alpha and beta funds. In fact, today most asset managers offer “portable alpha” products, and a large proportion of institutional investors have invested in these products. The returns on these alpha portfolios are often referred to as “absolute returns” because they are supposed to remain positive under all market conditions.

⁴The analysis for a large number of assets is not just an abstract mathematical device but also corresponds to practice: hedge funds and sovereign-wealth funds hold a large number of assets in their portfolios; for instance, the portfolio of Norges Bank has over 9,000 assets. Moreover, our results bite even when the number of assets is as small as 100.

mean-variance portfolios. We demonstrate using simulations in Section 5 how these results can be applied to improve the estimation of portfolio weights that achieve superior out-of-sample performance. We conclude in Section 6. Proofs and technical details for all our results are collected in the appendix.

2 Related Literature

In this section, we discuss the literature that is related to our work, in particular the area of robust control. We also discuss the literature related to the APT and to portfolio choice in the presence of estimation error.

Savage (1954) describes axioms under which a rational person can express her uncertainty in terms of a unique prior. In contrast, Gilboa and Schmeidler (1989) relax one of Savage’s axioms to axiomatize a class of preferences that capture ambiguity; these preferences are used in Chen and Epstein (2002) and Epstein and Wang (1994) and are called multiple-priors preferences. As the name suggests, a key feature of these preferences is that investors may have multiple priors about the future states of the world. An alternative approach assumes that investors allow for the possibility that the model they are using may not be the correct one and hence consider deviations from the reference model, where the likelihood of the approximating model relative to the reference model is measured using relative entropy; these preferences are called multiplier preferences and are discussed in Hansen and Sargent (2008). Maccheroni, Marinacci, and Rustichini (2006) show formally the relation between multiple-priors preferences and multiplier preferences.⁵ The approach that we adopt is similar to that described in Hansen and Sargent (2008).⁶

Our robust portfolio choice model is built on the APT, in the sense that the APT provides the framework for identifying the *approximating model* of asset returns around which the robust max-min decision is made. Ross (1976, 1977) develops the APT by showing that

⁵Maccheroni, Marinacci, and Rustichini (2006) also show that both classes of preferences are nested in a new class of preferences called “divergence preferences,” which are smooth, in contrast to preferences with multiple priors. Klibanoff, Marinacci, and Mukerji (2005) also develop a model of smooth ambiguity in which the investor uses a nonlinear function to evaluate expected utility values. All these models are based on a violation of the independence axiom. Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2011) provides a common representation that unifies these preferences without requiring independence over uncertain acts. For surveys of this literature, see Hansen and Sargent (2008), Guidolin and Rinaldi (2009), Epstein and Schneider (2010), and Gilboa and Marinacci (2016).

⁶The work described above abstracts from estimation issues; for one approach to the identification and estimation of dynamic models with robust decision makers, see Christensen (2017).

if asset returns have a strict factor structure (i.e., idiosyncratic components of returns are uncorrelated across assets), then mean returns are *approximately* linear functions of factor loadings. Huberman (1982) formalizes the argument proposed in Ross (1976). Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984) extend the results of Ross (1976, 1977) and Huberman (1982) to a setting where returns need to satisfy only an approximate factor structure; that is, the idiosyncratic components of returns are allowed to be mildly correlated across assets.

Just as in Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984), we do not restrict the covariance matrix of the residuals to be diagonal; that is, we allow for correlated error terms. However, in contrast to Chamberlain (1983) and Chamberlain and Rothschild (1983), we allow for large pricing errors, for instance, when they are related to some latent pervasive factors. More formally, we consider the case where the maximum eigenvalue of the residual covariance matrix is not restricted to be bounded as the number of assets increases. A statistical criterion to assess whether the error terms in a given model share at least one common (pervasive) factor is provided by Gagliardini, Ossola, and Scaillet (2017).

All these models deal with a large but countable number of assets. Building on the work of Al-Najjar (1998), Gagliardini, Ossola, and Scaillet (2016) extend the APT to allow for an uncountable number of assets and also relax the boundedness assumption of the maximum eigenvalue of the residual covariance matrix. In particular, Gagliardini, Ossola, and Scaillet (2016) show that the APT bound-inequality leads to zero pricing error for each asset when there is a continuum of assets. This is partly a consequence of the fact that they consider the *unweighted* sum of the squared pricing errors, as in the traditional APT setting. In the same setting with a continuum of assets, Renault, van der Heijden, and Werker (2017) extend the APT to squared excess returns allowing one to price common factors in the idiosyncratic variance of returns.

It is well known that mean-variance efficient portfolios that are based on sample estimates of first and second moments perform poorly out of sample; see, for example, Jobson and Korkie (1980), Frost and Savarino (1986), and DeMiguel, Garlappi, and Uppal (2009). Of the many approaches considered to improve the out-of-sample performance of mean-variance portfolios, one is to impose portfolio constraints; Jagannathan and Ma (2003) explain that the reason for the improved performance is that imposing shortsale constraints

is equivalent to shrinking the covariance matrix. Jorion (1986), Pástor (2000), and Pástor and Stambaugh (2000) show how shrinkage can arise in the context of Bayesian estimation methods. DeMiguel, Garlappi, Nogales, and Uppal (2009) show that further gains are possible by imposing a more general form of the shortsale constraint, a norm constraint on the portfolio weights, with this constraint also having the interpretation of Bayesian shrinkage. Pettenuzzo, Timmermann, and Valkanov (2014) demonstrate that economic constraints, such as restricting the equity risk premium to be positive and bounding the Sharpe ratio, improve the estimation of time-series forecasts of the equity risk premium. In our work, the constraint to be imposed when estimating asset returns follows naturally from the extended APT and the constraint imposed in the context of portfolio selection follows from the preference for robustness; moreover, the APT constraint can be interpreted as a constraint on the Sharpe ratio of the “alpha” portfolio.

Estimating the first and especially the second moments of asset returns in order to compute the optimal mean-variance portfolio weights suffers from the curse of dimensionality: if there are N risky assets, then the number of parameters to be estimated is of the order N^2 . Brandt (1999) proposes a *nonparametric* approach to estimate the portfolio weights directly from the Euler first-order conditions that depend on a given set of predictors, bypassing entirely the estimation of first and second moments. Ait-Sahalia and Brandt (2001) point out that this nonparametric approach cannot handle a large number of predictors and propose a *semiparametric* approach, where the portfolio weights depend nonparametrically on a single linear combination of the predictors. Brandt, Sant-Clara, and Valkanov (2009) provide a *parametric* approach for estimating the portfolio weights directly. In particular, they specify a parsimonious K -factor structure for the portfolio weights, where the number of free parameters K is less than the number of assets N and does not increase with N . Our work shares many analogies with these papers. Just like Brandt, Sant-Clara, and Valkanov (2009), we also rely on a factor model, but one that is for returns, in the sense of the extended APT, instead of portfolio weights. The advantage of doing this is that it allows us to tackle not just the problem of estimation error, thanks to the parsimony of the factor structure, but also the issue of misspecification, such as missing or mismeasured factors, by allowing for the possibility of non-zero pricing errors. Although this implies that the number of free parameters will increase (linearly) with N , we show precisely how by allowing for a large number of assets, misspecification in the risk premia and covariance of the observed factors can be fully resolved.

Finally, several papers in the robust control paper use the max-min approach to study the portfolio-choice problem. These papers include Dow and Werlang (1992), Trojani and Vanini (2002), Uppal and Wang (2003), Garlappi, Uppal, and Wang (2007), and Guidolin and Rinaldi (2009). In contrast to these papers, our work adopts the extended APT model as the *approximating model* and shows that allowing for pricing errors in a factor model results in both new theoretical insights and substantial improvement in the empirical out-of-sample performance. The insights we develop go beyond the “max-min” approach and could also be applied to other approaches used to model ambiguity.

3 Generalizing the APT

In this section, we first state the APT result in the existing literature, and then show how it can be extended to incorporate large pricing errors that are related to missing pervasive factors, and conclude by explaining how we use the APT to define the *set* of possibly misspecified models that contains the true asset-pricing model.

3.1 Arbitrage Pricing Theory (APT)

The number of risky assets is denoted by N . Just like in Chamberlain and Rothschild (1983) and Ingersoll (1984), we study a market with an infinite number of assets.⁷ Let r_f denote the return on the risk-free asset and let the N -dimensional vector $\mathbf{r}_t = (r_{1t}, r_{2t}, \dots, r_{Nt})'$ denote the vector of returns on risky assets. Given an arbitrary portfolio strategy a with weights $\mathbf{w}_N^a = (w_1^a, w_2^a, \dots, w_N^a)'$ of N risky assets, and using $\mathbf{1}_N$ to denote an N -dimensional vector of ones, we define the associated portfolio return as

$$r_t^a = \mathbf{r}_t' \mathbf{w}_N^a + r_f(1 - \mathbf{1}_N' \mathbf{w}_N^a),$$

with finite conditional mean, standard deviation, and Sharpe ratio defined as

$$\begin{aligned} \mu^a &= E(r_t^a) = E(\mathbf{r}_t)' \mathbf{w}_N^a + r_f(1 - \mathbf{1}_N' \mathbf{w}_N^a), \\ \sigma^a &= \sqrt{\text{var}(r_t^a)}, \quad \text{and} \quad \text{SR}^a = \frac{\mu^a - r_f}{\sigma^a}. \end{aligned}$$

⁷To make clear the dependence on the number of assets, we index quantities that are N -dimensional by the subscript N , except for random variables, such as the returns on risky assets, which have the subscript t . Instead of considering a sequence of distinct economies, we consider a *fixed* infinite economy in which we study a sequence of nested subsets of assets. Therefore, in the N th step, as a new asset is added to the first $N - 1$ assets, the parameters of the first $N - 1$ stay unchanged. These unchanging parameters can be interpreted as the parameters one would get in the limit as the number of assets becomes asymptotically large.

We start our analysis with the assumption of a linear latent-factor structure for returns. Just as in Chamberlain (1983), Chamberlain and Rothschild (1983), and Ingersoll (1984), we do not restrict the covariance matrix of the residuals, Σ_N , to be diagonal.

Assumption 1 (Linear factor model). *We assume the N -dimensional vector \mathbf{r}_t of asset returns can be characterized by*

$$\mathbf{r}_t = \boldsymbol{\mu}_N + \mathbf{B}_N \mathbf{z}_t + \boldsymbol{\varepsilon}_t, \quad (1)$$

where $\boldsymbol{\mu}_N$ is the $N \times 1$ vector of expected returns and \mathbf{B}_N is the $N \times K$ full-rank matrix of factor loadings, with $K < N$. At any time t , the $K \times 1$ vector of common unobserved factors, \mathbf{z}_t , is distributed with zero mean and $K \times K$ covariance matrix $\boldsymbol{\Omega}$, and the $N \times 1$ vector of residuals $\boldsymbol{\varepsilon}_t$ is distributed with zero mean and the $N \times N$ covariance matrix Σ_N , with $\boldsymbol{\Omega}$ and Σ_N being positive definite. Moreover, $\boldsymbol{\varepsilon}_t$ and \mathbf{z}_t are uncorrelated.

Assumption 1 implies that the variance-covariance matrix for asset returns is⁸

$$E [(\mathbf{r}_t - \boldsymbol{\mu}_N)(\mathbf{r}_t - \boldsymbol{\mu}_N)'] = \mathbf{V}_N = \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}_N' + \Sigma_N.$$

In the assumption below, as well as throughout the paper, we use δ to denote an arbitrary positive scalar, not necessarily having the same value.

Assumption 2 (No asymptotic arbitrage). *There is no sequence of portfolios for which, along some subsequence N' :*

$$\text{var}(\mathbf{r}'_t \mathbf{w}_{N'}^a) \rightarrow 0 \quad \text{as } N' \rightarrow \infty \quad \text{and} \quad (\boldsymbol{\mu}_{N'} - r_f \mathbf{1}_{N'})' \mathbf{w}_{N'}^a \geq \delta > 0 \quad \text{for all } N'.$$

We now state the APT result. This result is derived in Huberman (1982) and Ingersoll (1984), so we state it (and other results derived elsewhere) as propositions without proof.

⁸It is important to note that the APT is a model of the random component of returns, $\mathbf{r}_t - \boldsymbol{\mu}_N = \mathbf{B}_N \mathbf{z}_t + \boldsymbol{\varepsilon}_t$, and is silent about expected returns. Black (1995, p. 168) recognizes this and states that the ‘‘Arbitrage Pricing Theory (APT) is a model of variance. It says that the number of independent factors influencing return is limited, but it is silent on the pricing of these factors, so it is silent on expected return.’’ Notice that in the original formulation of the APT in (1), no factor is assumed to be observed; instead, the error term $\mathbf{r}_t - \boldsymbol{\mu}_N = \mathbf{B}_N \mathbf{z}_t + \boldsymbol{\varepsilon}_t$ has a latent factor structure through the common innovation \mathbf{z}_t . In contrast, nowadays factor models are typically used assuming the existence of a given number of observed factors, which, for simplicity, are assumed to be traded. This means that rather than the model in (1), one considers $\mathbf{r}_t - r_f \mathbf{1}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N (\mathbf{f}_t - r_f \mathbf{1}_K) + \boldsymbol{\varepsilon}_t$, for the K -dimensional factors \mathbf{f}_t , all of which are observed. Observe that this equation and (1) are equivalent if $\mathbf{f}_t - r_f \mathbf{1}_K = \boldsymbol{\lambda} + \mathbf{z}_t$ and $\boldsymbol{\mu}_N - r_f \mathbf{1}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}$.

Proposition 1 (The APT restriction (Huberman, 1982; Ingersoll, 1984)). *Assumptions 1 and 2 imply that for all N there exists some positive number δ such that the weighted sum of the squared pricing errors is uniformly bounded:*

$$\check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N \leq \delta < \infty, \quad (2)$$

where the vector of true pricing errors is

$$\check{\alpha}_N = (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) - \mathbf{B}_N \check{\boldsymbol{\lambda}}, \quad (3)$$

and the vector of true risk premia is

$$\check{\boldsymbol{\lambda}} = (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N), \quad (4)$$

with the $\check{\cdot}$ symbol used to denote that the quantity is obtained from the projection based on population parameters.⁹

Observe that $\check{\boldsymbol{\lambda}}$ is the projection coefficient when projecting $(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$ on \mathbf{B}_N , and $\check{\alpha}_N$ is the projection residual (pricing error) that satisfies

$$\mathbf{B}'_N \Sigma_N^{-1} \check{\alpha}_N = \mathbf{0}.$$

3.2 Extending the APT

In this section, we extend the standard APT to the case the pricing errors are large and potentially related to (missing) factors so that even a well-spread portfolio may not be well diversified; i.e., the residual covariance matrix does not necessarily have bounded eigenvalues. This could occur precisely because of model misspecification.

We adopt the following notation. Consider a symmetric $M \times M$ matrix \mathbf{A} . Let $g_{iM}(\mathbf{A})$ denote the i th eigenvalue of \mathbf{A} in decreasing order for $1 \leq i \leq M$. Thus, the maximum eigenvalue is $g_{1M}(\mathbf{A})$ and the minimum eigenvalue is $g_{MM}(\mathbf{A})$.

Next, we introduce the limit of $\check{\boldsymbol{\lambda}}$ as $N \rightarrow \infty$, which we label $\boldsymbol{\lambda}$, and the associated vector of pricing errors, $\boldsymbol{\alpha}_N = ((\boldsymbol{\mu}_N - r_f \mathbf{1}_N) - \mathbf{B}_N \boldsymbol{\lambda})$.¹⁰ If $g_{1K}((\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1}) \rightarrow 0$ as

⁹Note that $\check{\alpha}_N$ has the subscript N because it is N dimensional, while $\check{\boldsymbol{\lambda}}$ is not subscripted by N because it is not N dimensional, although it varies with N .

¹⁰Note that $\boldsymbol{\alpha}_N$ is an N -dimensional vector that is a component of the infinite-dimensional vector $\boldsymbol{\alpha}$; as N increases, the number of elements in $\boldsymbol{\alpha}_N$ increases, but the elements themselves do not change, unlike $\check{\alpha}_N$.

$N \rightarrow \infty$ and under Assumptions 1 and 2, Ingersoll (1984, Theorem 3 and ftn. 10) shows that $\boldsymbol{\lambda}$ is unique and prices assets with bounded squared error:¹¹

$$\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty. \quad (5)$$

In the rest of the paper, it will be more convenient to express our results in terms of $\boldsymbol{\alpha}_N$ than $\tilde{\boldsymbol{\alpha}}_N$.

The restriction in (5), which is a consequence of asymptotic no arbitrage, links the pricing error $\boldsymbol{\alpha}_N$ to the residual covariance matrix, $\boldsymbol{\Sigma}_N$. There are two possible, complementary, cases for $\boldsymbol{\Sigma}_N$ as $N \rightarrow \infty$; we examine both. In the first case, pricing errors are unrelated to pervasive factors and therefore small on average (i.e., all the eigenvalues of $\boldsymbol{\Sigma}_N$ are *bounded*). In the second case, the pricing errors are related to pervasive factors, and therefore, can be large on average (i.e., at least one of the eigenvalues is *unbounded*). The existing APT literature has focused on studying the case in which the pricing errors are small (implying that all the eigenvalues of $\boldsymbol{\Sigma}_N$ are bounded), which—for completeness—is restated in the theorem below, with the proof for this provided in Huberman (1982) and Ingersoll (1984).

Proposition 2 (Constraint imposed on $\boldsymbol{\alpha}_N$ by no arbitrage for case with small pricing errors; Huberman (1982); Ingersoll (1984)). *If the pricing errors are small (i.e., $\boldsymbol{\Sigma}_N$ has bounded eigenvalues for large N), then the restriction in (5) requires the elements of the pricing-error vector $\boldsymbol{\alpha}_N$ to become small for large N in the following sense:*

$$\boldsymbol{\alpha}'_N \boldsymbol{\alpha}_N \leq g_{1N}(\boldsymbol{\Sigma}_N)(\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N) \leq \delta < \infty,$$

but without $\boldsymbol{\alpha}_N$ being tied down to $\boldsymbol{\Sigma}_N$.

We now show that the APT model in the existing literature can be extended to the case where, as N increases, some of the eigenvalues of $\boldsymbol{\Sigma}_N$ are *not* bounded, which typically occurs when there are missing or mismeasured factors leading to large pricing errors.

Theorem 1 (Constraint imposed on $\boldsymbol{\alpha}_N$ by no arbitrage for case with large pricing errors). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumptions 1 and 2. Suppose that for some finite $1 \leq p < N$ the following three conditions hold: (i) $\sup_N g_{pN}(\boldsymbol{\Sigma}_N) = \infty$;*

¹¹Observe that the APT restriction in (2) is expressed in terms of $\tilde{\boldsymbol{\alpha}}_N$, while the restriction in (5) is expressed in terms of $\boldsymbol{\alpha}_N$; in Lemma C4, we show the equivalence between these two conditions.

(ii) $\sup_N g_{p+1N}(\boldsymbol{\Sigma}_N) \leq \delta < \infty$; and, (iii) $\inf_N g_{NN}(\boldsymbol{\Sigma}_N) \geq \delta > 0$. Then, the APT restriction in (5) is satisfied by the pricing error $\boldsymbol{\alpha}_N$, represented as

$$\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N, \quad (6)$$

where \mathbf{A}_N is an $N \times p$ matrix whose j th column equals $g_{jN}^{\frac{1}{2}}(\boldsymbol{\Sigma}_N) \mathbf{v}_{jN}(\boldsymbol{\Sigma}_N)$, where $1 \leq j \leq p$, $\mathbf{v}_{jN}(\boldsymbol{\Sigma}_N)$ is the eigenvector of $\boldsymbol{\Sigma}_N$ associated with the eigenvalue $g_{jN}(\boldsymbol{\Sigma}_N)$, $\boldsymbol{\lambda}_{miss}$ is some $p \times 1$ vector, and \mathbf{a}_N is some non-zero $N \times 1$ vector that satisfies $\mathbf{a}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{a}_N \leq \delta < \infty$.

One can interpret the two components of $\boldsymbol{\alpha}_N$ in (6) in a variety of ways. The first term, $\mathbf{A}_N \boldsymbol{\lambda}_{miss}$, could be associated with p latent or missing pervasive factors, where \mathbf{A}_N are the factor loadings and $\boldsymbol{\lambda}_{miss}$ are the risk premia for the missing factors.¹² The second term, \mathbf{a}_N , is the part of the pricing error $\boldsymbol{\alpha}_N$ that is unrelated to pervasive factors; for instance, \mathbf{a}_N could be interpreted as representing managerial skills or views of analysts.

Theorem 1 shows that the common perception that the pricing error $\boldsymbol{\alpha}_N$ needs to be small in the APT is not accurate. In particular, if the maximum eigenvalue of the residual covariance matrix is not bounded, then the pricing errors can also be large *without* violating the APT restriction given in (5); this is illustrated in Figure 1. What Theorem 1 states is that if the pricing errors are large so that the maximum eigenvalue of $\boldsymbol{\Sigma}_N$ is asymptotically unbounded, then the contribution of the pricing error to the portfolio return could also be large, but for this to satisfy the no-arbitrage condition, any portfolio earning this high return would *not* be well diversified and would be bearing idiosyncratic risk.¹³

Ingersoll (1984, p. 1026) defines the setting of Theorem 2 as one with *bounded residual variation*; we label the setting of Theorem 1 as one with *unbounded residual variation*.

¹²Pervasiveness of the latent missing factors with loadings \mathbf{A}_N follows from the assumption that g_{jN} for $1 \leq j \leq p$ are diverging for large N .

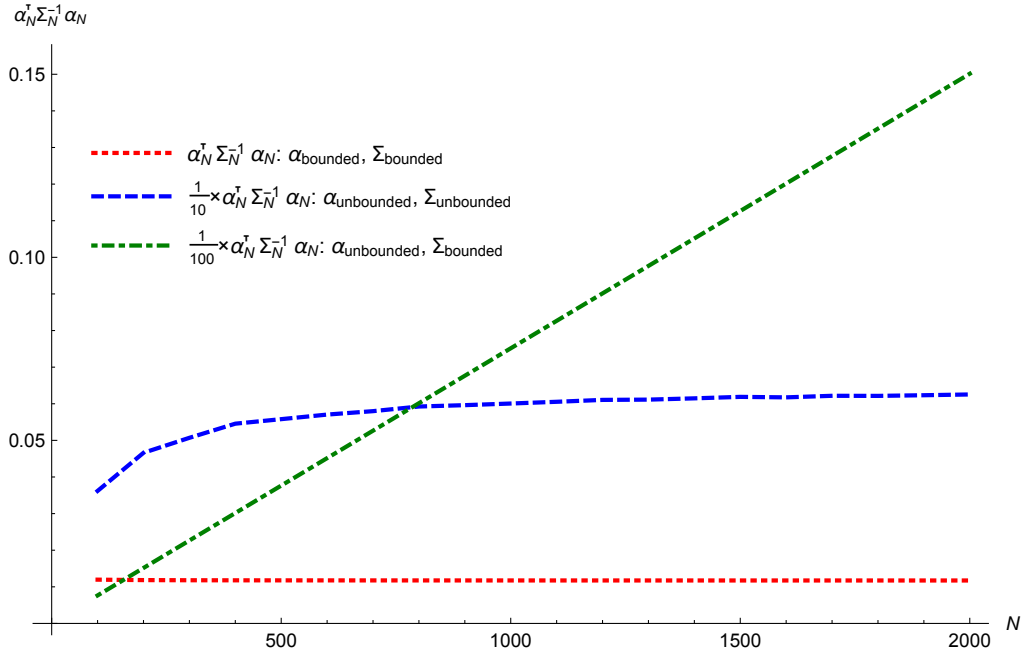
¹³To see this, observe that by Chamberlain and Rothschild (1983, Theorem 4), under the assumptions made for deriving Theorem 1, the covariance matrix of residuals, $\boldsymbol{\Sigma}_N$, has the following approximate p -factor structure: $\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$, where \mathbf{C}_N is a $N \times N$ positive semi-definite matrix with bounded eigenvalues. Under the assumptions of Chamberlain and Rothschild (1983, Theorem 4), \mathbf{C}_N will be non-singular for a sufficiently large N , which implies that $\boldsymbol{\Sigma}_N$ is also nonsingular for a sufficiently large N . Therefore, the residual variance of the return on any portfolio weights \mathbf{w}_N is given by

$$\mathbf{w}'_N \boldsymbol{\Sigma}_N \mathbf{w}_N = \mathbf{w}'_N \mathbf{A}_N \mathbf{A}'_N \mathbf{w}_N + \mathbf{w}'_N \mathbf{C}_N \mathbf{w}_N.$$

Whereas the second term, $\mathbf{w}'_N \mathbf{C}_N \mathbf{w}_N$, goes to zero for well-spread portfolios (that is, for portfolios with $\mathbf{w}'_N \mathbf{w}_N \rightarrow 0$), there is no guarantee that the same occurs for the first term, $\mathbf{w}'_N \mathbf{A}_N \mathbf{A}'_N \mathbf{w}_N$.

Figure 1: Weighted sum of the squared pricing errors

In this figure, we plot the weighted sum of the squared pricing errors, $\alpha_N^T \Sigma_N^{-1} \alpha_N$, as the number of assets N increases for three different cases. In the first case, which is the one studied in the existing literature, the elements of the pricing-error vector, α_N , become small as N increases and the eigenvalues of Σ_N are bounded; in this case, the weighted sum of the squared pricing errors is bounded, as shown by the dotted (red) line. In the second case, we allow for large pricing errors that are related to factors. In this case, some of the eigenvalues of Σ_N are *unbounded*. As the dashed (blue) line shows, even in this case the weighted sum of the squared pricing errors is bounded, demonstrating that the pricing errors can be large without violating the APT restriction. The third case, illustrated by the dotted-dashed (green) line, shows that if the pricing errors were large but the eigenvalues of Σ_N were *incorrectly* specified to be bounded, only then would the APT restriction be violated.



3.3 Constructing the Approximating Model

The typical interpretation of the inequality in (5) is that the *true* asset-pricing model characterized by the pricing errors α_N rules out arbitrage. In addition, we interpret this inequality as defining the *set* of models that nests the approximating model, which is the estimated model within this set that best describes the data. The approximating model plays a central role in determining the performance of the robust portfolio weights, and hence, it is important to specify and estimate this model accurately. In contrast, as Hansen

and Sargent (2008, Section 1.7) state, most of the literature on robust control is “silent about the process through which the decision maker discovered her approximating model.”

To determine the approximating model, the key quantity to be estimated is α_N . Although virtually any estimation procedure could be used, we will illustrate it with respect to the (pseudo) Gaussian maximum-likelihood (ML) estimator. As explained by Hansen and Sargent, ML estimation is desirable; in our case, it is even more suitable because it allows us to include the constraint in (5) in a natural way when estimating the APT model. Moreover, as we explain below, allowing for the presence of pricing errors mitigates the fragility typically associated with ML estimation in the presence of model misspecification, for example, because of the possibility of missing factors.

We propose a multi-step procedure to determine in which case we are. In the first step, one estimates the parameters of the factor model conditional on the factor realizations without imposing the APT restriction. Having obtained consistent estimates of the parameters, the next step is to analyze the possibility of pervasive missing factors by studying the eigenvalues associated with the estimated $\hat{\Sigma}_N$, where the $\hat{\cdot}$ denotes an estimated quantity. This part uses conventional principal-component analysis of $\hat{\Sigma}_N$, and it allows one to estimate the number of latent pervasive factors, p ; see, for example, Anderson (1984) and, more recently, the procedure in Gagliardini, Ossola, and Scaillet (2017).

In the next two subsections, we provide the details of how to estimate the model when $\hat{p} = 0$ (i.e., small pricing-error case) and when $\hat{p} > 0$ (i.e., large pricing-error case), after imposing the APT constraint. In both cases, we obtain the estimated pricing errors, $\hat{\alpha}_N$, which plays a central role in the derivation of the optimal portfolio weights that are robust to model misspecification.

3.3.1 Estimation for the case of small pricing errors

In the case where the pricing errors are small (that is, $\hat{p} = 0$ implying that $\alpha_N = \mathbf{a}_N$), under Assumption 1, the true unconditional means and covariances of returns satisfy

$$E(\mathbf{r}_t - r_f \mathbf{1}_N) = \boldsymbol{\mu}_N - r_f \mathbf{1}_N = \boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}, \quad \text{var}(\mathbf{r}_t) = \mathbf{V}_N = \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}_N' + \boldsymbol{\Sigma}_N,$$

where $\boldsymbol{\lambda} = E(\mathbf{f}_t) - r_f \mathbf{1}_K$ is the vector of risk premia, $\boldsymbol{\Omega} = \text{var}(\mathbf{f}_t)$ is the covariance matrix of the factors assuming, without loss of generality, stationarity of the K factors, \mathbf{f}_t , and

that the factors are traded. The more general case of estimation with both traded and non-traded factors is described in the appendix.

The (pseudo) Gaussian ML estimator for $\boldsymbol{\theta} = (\boldsymbol{\alpha}'_N, \text{vec}(\mathbf{B}_N)', \text{vech}(\boldsymbol{\Sigma}_N)', \boldsymbol{\lambda}', \text{vech}(\boldsymbol{\Omega})')'$, under the assumption that the data is independently and identically distributed, coincides with the OLS estimator for $\boldsymbol{\alpha}_N, \mathbf{B}_N, \boldsymbol{\Sigma}_N$, conditional on the realization of the factors; moreover, $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$ coincide with the sample mean and sample covariance of the factors, \mathbf{f}_t .¹⁴ However, because the APT restriction is not guaranteed to hold, one should consider the ML-*constrained* (MLC) estimator:

$$\hat{\boldsymbol{\theta}}_{\text{MLC}} = \underset{\tilde{\boldsymbol{\theta}}}{\text{argmax}} L(\tilde{\boldsymbol{\theta}}) \quad \text{subject to} \quad \tilde{\boldsymbol{\alpha}}'_N \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\boldsymbol{\alpha}}_N \leq \delta, \quad (7)$$

where, based on a sample of size T , $L(\tilde{\boldsymbol{\theta}})$ is the Gaussian log-likelihood function defined in equation (B1) of the appendix and $\tilde{\boldsymbol{\theta}}$ defines the set of feasible parameter values.¹⁵

Because the parameter $\boldsymbol{\alpha}_N$ is now constrained by the APT restriction, imposing this constraint should lead $\hat{\boldsymbol{\theta}}_{\text{MLC}}$ to be a more precise estimator of the true parameter values compared to the unconstrained estimator. As we show in Theorem B1 in the appendix, the MLC estimator for $\boldsymbol{\alpha}_N$ is

$$\hat{\boldsymbol{\alpha}}_{N,\text{MLC}} = \frac{1}{1 + \hat{\kappa}} \left[\bar{\mathbf{r}} - r_f \mathbf{1}_N - \hat{\mathbf{B}}_{N,\text{MLC}} (\bar{\mathbf{f}} - r_f \mathbf{1}_K) \right],$$

where $\hat{\kappa}$ is the Karush-Kuhn-Tucker multiplier for the inequality constraint, $\bar{\mathbf{r}}$ and $\bar{\mathbf{f}}$ are the sample means of the returns and the factors, respectively, and $\hat{\mathbf{B}}_{N,\text{MLC}}$ is the MLC estimator for \mathbf{B}_N . The constrained estimator $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$ turns out to be precisely the ridge estimator for $\boldsymbol{\alpha}_N$. However, without loss of generality, we assume that the δ in the APT constraint is large enough so that, at the true parameter values, $\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N$ is strictly less than δ . In turn, this implies that the Gaussian MLE $\hat{\boldsymbol{\alpha}}_N$ has the conventional limit distribution with covariance matrix $(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}) \boldsymbol{\Sigma}_N$ because “boundary issues” are avoided.¹⁶

¹⁴Note that $\text{vec}(\cdot)$ denotes the operator that stacks the columns of a matrix into a single column vector, and $\text{vech}(\cdot)$ denotes the operator that stacks the unique elements of the columns of a symmetric matrix into a single column vector.

¹⁵The APT theory does not tell us what δ should be in (7) and also in (8) below. In our estimation of the model, we use cross validation to identify δ .

¹⁶Typically, when the true parameter value lies on the boundary of the parameter space, the asymptotic distribution of the MLE is non-standard. In particular, Andrews (1999) shows that such asymptotic distribution can be expressed by the distribution of a random vector that minimizes a stochastic quadratic function over a convex cone, function of the modified parameter space. Often such distribution can be simulated, and depends on nuisance parameters that can be estimated consistently.

3.3.2 Estimation for the case of large pricing errors

The second setting in which alpha misspecification arises is when there are $\hat{p} > 0$ missing pervasive factors. Observe that the APT restriction is always satisfied for the case of *only* missing pervasive factors (that is, the case in which $\mathbf{a}_N = \mathbf{0}$), once we recognize that Σ_N contains the loadings of the missing factors, \mathbf{A}_N ;¹⁷ MacKinlay and Pástor (2000) use this insight to improve the precision of the estimated \mathbf{A}_N parameters. However, for the general case with large pricing errors (i.e., the unbounded residual-variation case), where $\alpha_N = \mathbf{A}_N \lambda_{miss} + \mathbf{a}_N$, the APT restriction is not automatically satisfied because \mathbf{a}_N is not in Σ_N . Therefore, when estimating the model, we need to impose the constraint, $\mathbf{a}'_N \mathbf{a}_N \leq \delta < \infty$ for any N ; that is:

$$\hat{\theta}_{MLC} = \underset{\tilde{\theta}}{\operatorname{argmax}} L(\tilde{\theta}) \quad \text{subject to} \quad \tilde{\mathbf{a}}'_N \tilde{\Sigma}_N^{-1} \tilde{\mathbf{a}}_N \leq \delta, \quad (8)$$

where the log-likelihood function $L(\tilde{\theta})$ is now defined in (B2), and where $\tilde{\theta}$ defines the set of feasible parameter values $\tilde{\theta} = (\tilde{\mathbf{a}}'_N, \tilde{\lambda}_{miss}, \operatorname{vec}(\tilde{\mathbf{A}}_N)', \operatorname{vec}(\tilde{\mathbf{B}}_N)', \operatorname{vech}(\tilde{\mathbf{C}}_N)', \tilde{\lambda}', \operatorname{vech}(\tilde{\Omega}')'$. As shown in Theorem B2, when the Karush-Kuhn-Tucker multiplier for the inequality constraint $\hat{\kappa}$ equals zero, one can identify only $\alpha_N = \mathbf{A}_N \lambda_{miss0} + \mathbf{a}_N$, but not the two components separately. In contrast, when $\hat{\kappa} > 0$, the APT restriction not only leads to shrinkage of the estimator for \mathbf{a}_N , but represents also the identification condition required to estimate λ_{miss} and \mathbf{a}_N separately.¹⁸ Finally, as we show in the appendix, it is remarkable that the ML estimate of λ_{miss} turns out to be exactly what one would obtain from the classical two-pass generalized-least-square (GLS) estimator of risk premia for nontraded factors Cochrane (2005). Therefore, this is another important motivation for using the ML approach: the non-observability of the missing factors make them necessarily nontraded and ML estimation automatically takes this into account.

¹⁷To see this, consider the case where $\alpha_N = \mathbf{A}_N \lambda_{miss}$ and $\Sigma_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$, where λ_{miss} is the risk premia corresponding to the missing factors, and \mathbf{C}_N is an $N \times N$ positive-definite matrix with bounded eigenvalues. Then, using the Sherman-Morrison-Woodbury formula, it follows that $\alpha'_N \Sigma_N^{-1} \alpha_N = \lambda'_{miss} \mathbf{A}'_N \Sigma_N^{-1} \mathbf{A}_N \lambda_{miss} = \lambda'_{miss} (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N) \lambda_{miss}$. Thus, $\alpha'_N \Sigma_N^{-1} \alpha_N$ converges to $\lambda'_{miss} \lambda_{miss}$ as $N \rightarrow \infty$ because $(\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ converges to the identity matrix, given that the missing factors are pervasive, implying that $(\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ is increasing without bound.

¹⁸Note that \mathbf{A}_N can only be estimated up to an orthogonal rotation and the same applies to λ_{miss} (when $\hat{\kappa} > 0$).

4 The Robust Portfolio Choice Problem

In the section above, we have shown how to estimate the approximating model that the investor believes is closest to the true model, represented by the estimates pricing errors, $\hat{\boldsymbol{\alpha}}_N$. In this section, we derive the optimal portfolio of an investor who recognizes that the approximating model may not only be misspecified but itself is certainly affected by estimation error. In order to make decisions that are robust, the investor chooses a max-min approach, which we describe below.

4.1 The Max-Min Portfolio Problem

The max-min approach is a useful framework to achieve a robust decision rule because it provides a lower bound on the performance over a range of alternative models. Gilboa and Schmeidler (1989) formalize the max-min approach in terms of a rational decision maker who has multiple priors and is assuming that the environment that she is confronted with (i.e., nature) will minimize her expected utility.

Theorem 2 (Robust portfolio weights). *The solution, \mathbf{w}_N^{rmv} , to the robust mean-variance (rmv) portfolio choice problem of an investor with risk aversion $\gamma > 0$ is*

$$\max_{\mathbf{w}_N} \min_{\boldsymbol{\alpha}_N} \left\{ \mathbf{w}'_N (\boldsymbol{\mu}_N - \mathbf{1}_N r_f) - \frac{\gamma}{2} \mathbf{w}_N \mathbf{V}_N \mathbf{w}_N \right\},$$

subject to the relative entropy constraint

$$\int_{-\infty}^{\infty} \left(\ln \frac{f_{\boldsymbol{\alpha}}(\mathbf{r}_t^e)}{f_{\hat{\boldsymbol{\alpha}}}(\mathbf{r}_t^e)} \right) f_{\boldsymbol{\alpha}}(\mathbf{r}_t^e) d\mathbf{r}_t^e \leq \delta,$$

where $\mathbf{r}_t^e = \mathbf{r}_t - r_f \mathbf{1}_N$ is assumed to satisfy $f_{\boldsymbol{\alpha}}(\mathbf{r}_t^e) \sim \mathcal{N}(\boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}, \mathbf{V}_N)$, is

$$\mathbf{w}_N^{rmv} = \frac{1}{\gamma} (\phi \boldsymbol{\Sigma}_N + \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N)^{-1} (\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda}), \quad (9)$$

where

$$\phi = \left[1 + \frac{\sqrt{2}}{\gamma} \left(\frac{\delta}{\mathbf{w}_N^{rmv} \boldsymbol{\Sigma}_N \mathbf{w}_N^{rmv}} \right)^{1/2} \right]. \quad (10)$$

Observe that under the assumptions of Theorem 2, the relative entropy constraint simplifies to

$$\frac{1}{2} (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)' \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) \leq \delta. \quad (11)$$

Therefore, a larger δ represents an increase in the investor's degree of aversion to model misspecification. We show in the appendix that if one sets $\delta = \frac{1}{2} \frac{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{T} \chi_{N,x\%}^2$, where $\chi_{N,x\%}^2$ corresponds to the x^{th} ($0 \leq x \leq 1$) quantile of a χ_N^2 distribution, then the entropy constraint defines the set of $\boldsymbol{\alpha}_N$ that are statistically indistinguishable from $\hat{\boldsymbol{\alpha}}_N$ at $(1-x)\%$, namely:

$$T(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)' \frac{\boldsymbol{\Sigma}_N^{-1}}{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})} (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) \leq \chi_{N,x\%}^2 < \infty.$$

More importantly, in this circumstance,

$$\phi = \left[1 + \frac{\sqrt{\chi_{N,x\%}^2}}{\gamma \sqrt{T}} \left(\frac{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{\mathbf{w}_N^{rmv}' \boldsymbol{\Sigma}_N \mathbf{w}_N^{rmv}} \right)^{1/2} \right]. \quad (12)$$

Note that although \mathbf{w}_N^{rmv} in (9) is defined in terms of ϕ , where ϕ in (10) is itself defined in terms of \mathbf{w}_N^{rmv} , one can obtain an explicit expression for \mathbf{w}_N^{rmv} by solving a fourth-order polynomial equation that, following Garlappi, Uppal, and Wang (2007), has a unique solution. For the case where N is large, which we discuss below, the fourth-order polynomial equation reduces to a linear equation. Moreover, one can show that ϕ is independent of risk aversion, γ .

Observe that the objective function entails both a maximization over the vector of portfolio weights, \mathbf{w}_N , which is standard in portfolio choice problems, and also a minimization over the pricing errors, $\boldsymbol{\alpha}_N$, which are constrained to lie in a given neighborhood of the estimated $\hat{\boldsymbol{\alpha}}_N$. Standard optimal control theory assumes that the decision maker knows the true model. In contrast, robust control theory treats the decision-maker's model as an approximation and seeks a single rule that works over a set of nearby models that might govern the data. The minimization over $\boldsymbol{\alpha}_N$ is a consequence of the investor's preference for robustness; the role of the constraint in (11) is to ensure that the chosen $\boldsymbol{\alpha}_N$ are statistically indistinguishable (for a given significance value) from the estimated $\hat{\boldsymbol{\alpha}}_N$. Using a set of perturbed models, that are difficult to distinguish statistically from the approximating model given the available data, protects the max-min decision maker from making decisions that are too conservative.¹⁹

¹⁹Observe that the max-min constraint under specification (12) uses the fact that the Gaussian MLE $\hat{\boldsymbol{\alpha}}_N$ has the conventional limit distribution with covariance matrix $(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}) \boldsymbol{\Sigma}_N$. If, moreover, the investor fears that the data are *not* i.i.d. normal, then the robust, sandwich-form, of the asymptotic covariance matrix for $\hat{\boldsymbol{\alpha}}_N$ should be used in the max-min constraint. In particular, $(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}) \boldsymbol{\Sigma}_N$ needs to be replaced by the corresponding block of $[E(\partial^2 \frac{L(\theta)}{\partial \theta \partial \theta'})]^{-1} E(\frac{\partial L(\theta)}{\partial \theta} \frac{\partial \partial L(\theta)}{\partial \theta'}) [E(\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta'})]^{-1}$, where $L(\theta)$ is the Gaussian loglikelihood, as discussed in Section 3.3. Details are available upon request.

Observe from (9) that in the absence of a concern for model misspecification, the expression for \mathbf{w}_N would have $\phi = 1$, which is the solution to the classical mean-variance portfolio problem. In contrast, the investor’s concern for misspecification implies that the residual variance, here denoted by Σ_N , is scaled up by $\phi > 1$. The magnitude of the scaling increases as we increase the size of the set of models over which the investor is uncertain; that is, increase δ .

A crucial observation is that the performance of any robust portfolio weights obtained from a max-min approach will depend critically on the quality of the approximating model because the minimization is performed around a small, statistically indistinguishable, perturbation of this approximating model. For instance, in Garlappi, Uppal, and Wang (2007), the approximating model is based on the sample mean and sample covariance of returns; these moments are notorious for being poorly estimated, especially when N is large. In contrast, our approximating model relies on the APT, whose merits are discussed above, and will be demonstrated below using empirical data.

The investor’s concern for misspecification leads to a scaling of the residual variance, Σ_N , rather than the overall variance, \mathbf{V}_N , because at this stage we are addressing misspecification in only the non-factor component of expected returns. If one were addressing misspecification in also the factor-component of expected returns, in particular the factor risk premia, then one would have scaled the variance of overall (residual plus factor component) returns. In our analysis, as we demonstrate below, it will be more effective to resolve misspecification in the factor risk premia using new insights that apply when the number of assets N is large.

We now show how the robust mean-variance portfolio weights can be written as a linear combination of two portfolios, each with special properties. The first portfolio, which we label the “alpha” portfolio, depends only on the pricing errors and the residual covariance matrix; the second portfolio, the “beta” portfolio, depends only on the moments of the common observed factors driving returns, namely, \mathbf{B}_N , $\mathbf{\Omega}$, and the risk premia, $\boldsymbol{\lambda}$.

Theorem 3 (Decomposing weights of robust portfolio). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumptions 1 and 2. Then as $N \rightarrow \infty$:*

(i) *The robust mean-variance portfolio weight satisfy the following decomposition:*

$$\mathbf{w}_N^{rmv} \sim \frac{1}{\gamma\phi} \mathbf{w}_N^\alpha + \frac{1}{\gamma} \mathbf{w}_N^\beta,$$

where the alpha and beta portfolios are, respectively:

$$\mathbf{w}_N^\alpha = \boldsymbol{\Sigma}_N^+ \hat{\boldsymbol{\alpha}}_N, \quad \mathbf{w}_N^\beta = \mathbf{V}_N^{-1} \mathbf{B}_N \boldsymbol{\lambda},$$

and we define

$$\boldsymbol{\Sigma}_N^+ = [\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1}], \quad (13)$$

where the symbol \sim denotes asymptotic equivalence.²⁰

(ii) For a sufficiently small δ , in particular such that $2\delta < (\hat{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^+ \hat{\boldsymbol{\alpha}}_N)$, the shrinkage parameter ϕ satisfies

$$\phi \sim \frac{1}{1 - \frac{(2\delta)^{\frac{1}{2}}}{(\hat{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^+ \hat{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}}}. \quad (14)$$

When $\delta = \frac{1}{2} \frac{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{T} \chi_{N,x\%}^2$, approximation (14) for ϕ becomes

$$\phi \sim \frac{1}{\left(1 - \left(\frac{\chi_{N,x\%}^2}{T}\right)^{\frac{1}{2}} \frac{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})^{\frac{1}{2}}}{(\mathbf{w}_N^\alpha \boldsymbol{\Sigma}_N^+ \mathbf{w}_N^\alpha)^{\frac{1}{2}}}\right)} \text{ as } N \rightarrow \infty. \quad (15)$$

To interpret the expression in (15), note that ϕ is increasing with $\chi_{N,x\%}^2$, which increases (approximately) linearly with the number of assets N , and $(\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})^{\frac{1}{2}}$, which is the Sharpe ratio of the beta portfolio as $N \rightarrow \infty$, and decreases with the sample size T (because the pricing errors are estimated more precisely) and $(\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^+ \boldsymbol{\alpha}_N)^{\frac{1}{2}}$, which is the Sharpe ratio of the alpha portfolio as $N \rightarrow \infty$.

It is useful to discuss the relation between \mathbf{V}_N^{-1} , the inverse of the covariance matrix of returns, and $\boldsymbol{\Sigma}_N^+$. Note that the Sherman-Morrison-Woodbury formula implies

$$\mathbf{V}_N^{-1} = [\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1}]. \quad (16)$$

Setting $\boldsymbol{\Omega}^{-1} = \mathbf{0}$ in (16) then leads to the expression for $\boldsymbol{\Sigma}_N^+$ in (13); that is, when $\boldsymbol{\Omega}$ (namely, its largest eigenvalue) takes arbitrarily large value, then \mathbf{V}_N^{-1} tends toward $\boldsymbol{\Sigma}_N^+$.²¹ To understand the intuition for this, observe that mean-variance optimization, that uses \mathbf{V}_N^{-1} , penalizes assets with large variances. As the elements of $\boldsymbol{\Omega}$ become arbitrarily large, all assets will have a large variance, and therefore, a low Sharpe ratio and expected quadratic

²⁰We say that $a_n \sim b_n$, if $a_n/b_n \rightarrow 1$ as $n \rightarrow \infty$.

²¹The expression in (16) also shows that \mathbf{V}_N^{-1} tends toward $\boldsymbol{\Sigma}_N^{-1}$ when the elements of $\boldsymbol{\Omega}$ (or, more precisely, its largest eigenvalue) take arbitrarily small values.

utility. The only way to avoid this outcome is to diversify the exposure of the portfolio to common factors; that is, the magnitude of the elements of $\mathbf{V}_N^{-1}\mathbf{B}_N$ to be as small as possible. This is automatically achieved when $\boldsymbol{\Omega}^{-1} = \mathbf{0}$ in the expression of \mathbf{V}_N^{-1} , which leads to $\boldsymbol{\Sigma}_N^+$, guaranteeing that $\boldsymbol{\Sigma}_N^+\mathbf{B}_N = \mathbf{0}$, which implies that the factor-related common risk is *perfectly* diversified by the \mathbf{w}_N^α portfolio for any N .

As illustrated in Figure 2, the weights of the \mathbf{w}_N^β portfolio are typically small and positive, while the weights of the \mathbf{w}_N^α portfolio are large and take both positive and negative values implying that the alpha portfolio has both long and short positions. In the next section, we show formally that as the number of assets increases, the weight of each asset in the beta portfolio is dominated by the corresponding weight in the alpha portfolio. This allows us to address beta misspecification, namely, to replace *without any loss in performance* the beta portfolio by a suitable benchmark portfolio (such as the equal-weighted or value-weighted portfolios) that is independent of any misspecification in the factor risk premia.

4.2 Mitigating misspecification in the beta component of returns

We now state our result about the properties of the alpha portfolio weights relative to the beta portfolio weights as the number of assets is asymptotically large.

Theorem 4 (Weights of alpha and beta portfolios for large N). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumptions 1 and 2 and $\boldsymbol{\alpha}_N \neq \mathbf{0}$. Suppose also that \mathbf{A}_N , \mathbf{B}_N , and $\mathbf{1}_N$ are \mathbf{C}_N -regular with the same scaling factor $f(N)$, and \mathbf{A}_N and \mathbf{B}_N are not asymptotically collinear.²² As $N \rightarrow \infty$, then:*

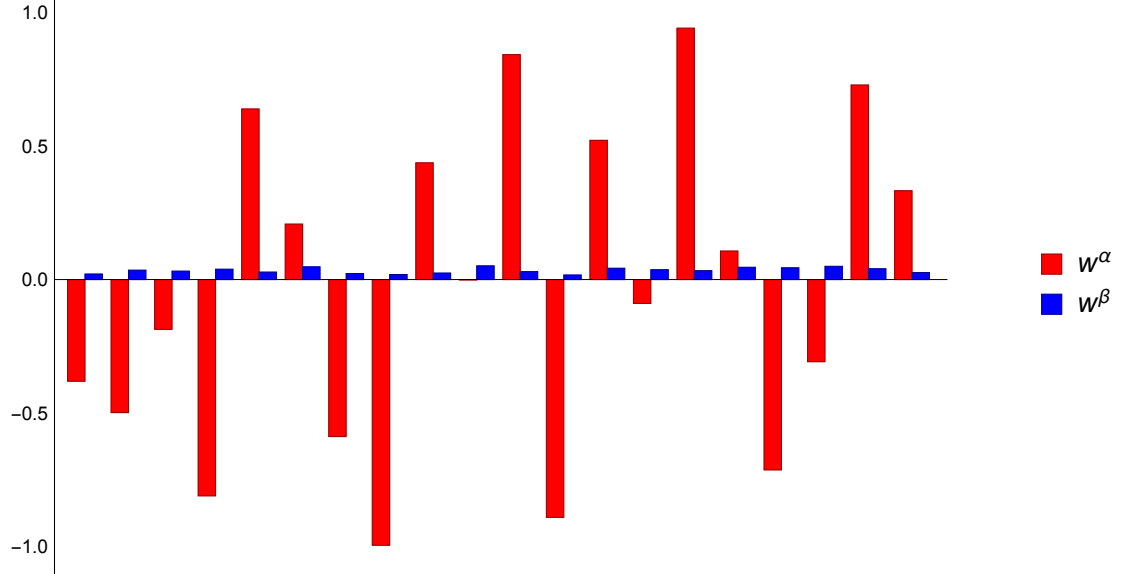
(i) *The elements of \mathbf{w}_N^α and \mathbf{w}_N^β converge to zero. Moreover, for every asset i and a constant $\delta_i \neq 0$, for the general case in which the pricing error unrelated to factors $\mathbf{a}_N \neq \mathbf{0}_N$*

$$\frac{w_{N,i}^\beta}{w_{N,i}^\alpha} \rightarrow 0, \quad (17)$$

²²Extending Ingersoll (1984, p. 1028), a matrix \mathbf{D}_N is \mathbf{C}_N -regular if there exists an increasing function of N , $f(N)$, such that for any $1 \leq j \leq K$, the eigenvalues $g_{jK}(\frac{1}{f(N)}\mathbf{D}'_N\mathbf{C}_N^{-1}\mathbf{D}_N) \rightarrow \delta_j > 0$, where δ_j is some finite positive constant. For example, in a factor economy with loading matrix \mathbf{D}_N and residual-covariance matrix \mathbf{C}_N , then \mathbf{C}_N -regularity holds whenever the factors are pervasive and there is no (pervasive) factor structure in the residual. By *asymptotic collinearity* we mean that either $\mathbf{A}'_N\mathbf{M}_{B_N}\mathbf{A}_N \rightarrow 0$ or $\mathbf{B}'_N\mathbf{M}_{A_N}\mathbf{B}_N \rightarrow 0$ or both, as N diverges, depending on whether the number of unobserved factors $p \leq K$, $p \geq K$ or $p = K$, where $\mathbf{M}_C = \mathbf{I}_N - \mathbf{C}_N(\mathbf{C}_N'\mathbf{C}_N)^{-1}\mathbf{C}_N'$ is the matrix that spans the space orthogonal to any full-column-rank matrix \mathbf{C}_N . When $p \leq K$, a sufficient condition for this is $\mathbf{A}_N = \mathbf{B}_N\boldsymbol{\delta} + \mathbf{G}_N$ for some constant $K \times p$ matrix $\boldsymbol{\delta}$ and some residual matrix \mathbf{G}_N satisfying $\mathbf{G}'_N\mathbf{G}_N \rightarrow 0$. When \mathbf{G}_N is a matrix of zeroes, then \mathbf{A}_N and \mathbf{B}_N are perfectly collinear.

Figure 2: Typical weights of the \mathbf{w}_N^α and \mathbf{w}_N^β portfolios

In this bar chart, we plot the typical weights of the \mathbf{w}_N^α (gray bars) and \mathbf{w}_N^β portfolios (black bars) for the case in which the number of assets is $N = 20$. The figure shows that the weights of the \mathbf{w}_N^β portfolio are small and positive. In contrast, the weights of the \mathbf{w}_N^α portfolio are large and take both positive and negative values.



whereas $w_{N,i}^\beta/w_{N,i}^\alpha \rightarrow \delta_i$ for the special case in which $\mathbf{a}_N = \mathbf{0}_N$.

(ii) The sum of the squared components of the mean-variance portfolio vectors $\mathbf{w}_N^\alpha{}' \mathbf{w}_N^\alpha$ is always bounded, whereas $\mathbf{w}_N^\beta{}' \mathbf{w}_N^\beta$ always converges to zero.

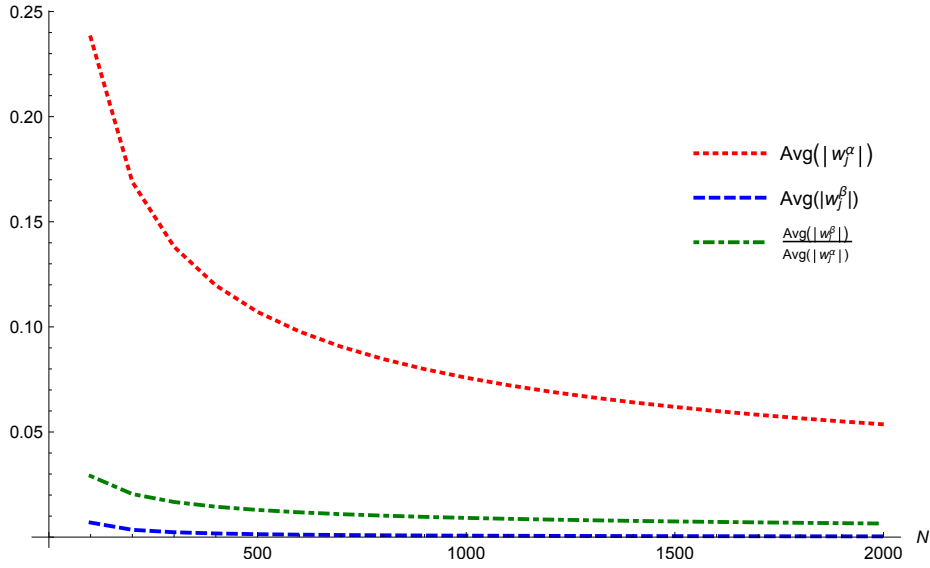
(iii) The sum of the components of the mean-variance portfolio vectors $|\mathbf{1}'_N \mathbf{w}_N^\alpha|$ can diverge to infinity, whereas $|\mathbf{1}'_N \mathbf{w}_N^\beta|$ is always bounded.

The \mathbf{w}_N^α portfolio dominates the \mathbf{w}_N^β portfolio across all three norms considered in the theorem above; this dominance is illustrated in Figure 3. The notion of diversification used in part (ii) of the theorem is the sum of the squares, which is the same notion adopted in Chamberlain (1983).²³ Part (iii) of the theorem studies how the total investment in

²³Because $\mathbf{w}_N^\beta{}' \mathbf{w}_N^\beta \geq \sup_i |w_{N,i}^\beta|$, it follows that the \mathbf{w}_N^β portfolio is diversified according to the sup norm criterion, which is the norm used in Green and Hollifield (1992). In contrast, the \mathbf{w}_N^α portfolio is not necessarily diversified according to the squared norm. For example, there could be a finite number of assets with a sufficiently large alpha, in which case the weights of these assets will not go to zero. Alternatively,

Figure 3: Relative average magnitude of weights of \mathbf{w}_N^α and \mathbf{w}_N^β portfolios

In this figure, we plot three quantities as the number of assets, N , increases. The three quantities are: (1) the average magnitude of the weights of the \mathbf{w}_N^α portfolio, given by the dotted (red) line at the top of the figure; (2) the average magnitude of the weights of the \mathbf{w}_N^β portfolio, given by the dashed (blue) line at the bottom of the figure; and (3) the ratio of the average magnitude of the weights of the \mathbf{w}_N^β portfolio to the corresponding weights of the \mathbf{w}_N^α portfolio, given by the dotted-dashed (green) line. The figure shows that as N increases, the average magnitude of the weights of the \mathbf{w}_N^β portfolio declines faster than the average magnitude of the weights of the \mathbf{w}_N^α portfolio.



risky assets is allocated between the \mathbf{w}_N^α and \mathbf{w}_N^β portfolios. The result in the theorem shows that $\mathbf{1}'_N \mathbf{w}_N^\alpha$ could be greater than 1 *and* it could be growing without bound as N increases, implying that it may be optimal to lever up unboundedly the investment in the \mathbf{w}_N^α portfolio.²⁴ On the other hand, the investment in the \mathbf{w}_N^β portfolio is bounded, and hence, is associated with a finite amount of leverage.

To understand the intuition for the results about the dominance of the \mathbf{w}_N^α portfolio weights, recall that the objective of mean-variance portfolio optimization is to maximize the portfolio Sharpe ratio, which entails increasing the mean of the portfolio return and/or

even if none of the assets has a particularly large alpha, the weights of the \mathbf{w}_N^α portfolio can go to zero at a sufficiently slow rate.

²⁴To see that $\mathbf{1}'_N \mathbf{w}_N^\alpha$ can diverge, consider the following example in which $\boldsymbol{\alpha}_N = \mathbf{1}_N / \sqrt{N}$, $\boldsymbol{\Sigma}_N = \sigma^2 \mathbf{I}_N$, with \mathbf{I}_N the $N \times N$ identity matrix, and there is a single factor with β_i with iid distribution having mean 1 and variance $\sigma_\beta^2 > 0$. Then, $\mathbf{1}'_N \mathbf{w}_N^\alpha \sim \sqrt{N} \sigma_\beta^2 / (1 + \sigma_\beta^2)$ which goes to infinity with N .

reducing the volatility of the portfolio return. There are two sources of risk: factor exposure and idiosyncratic exposure. The factor exposure of the \mathbf{w}_N^α portfolio is zero—*irrespective* of the rate at which the weights decrease—because of the orthogonality of \mathbf{w}_N^α to \mathbf{B}_N . Regarding exposure to idiosyncratic risk, the elements of \mathbf{w}_N^α cannot decrease faster than $1/N^{\frac{1}{2}}$ because then the idiosyncratic risk of the portfolio goes to zero; however, the idiosyncratic risk of the alpha portfolio coincides with its Sharpe ratio, implying that the Sharpe ratio would also go to zero. On the other hand, the APT restriction does not allow the rate at which the weights decrease to be slower than $1/N^{\frac{1}{2}}$. To understand this, consider the simple case in which Σ_N is the identity matrix. Recall that for the sum of N positive terms to be bounded, it suffices that each term declines not slower than $1/N$. The square-root rate follows from the fact that in our case each term is in fact the square of $w_{N,i}^\alpha$. Thus, the rate of $1/N^{\frac{1}{2}}$ strikes just the correct balance between optimizing the risk and return of the \mathbf{w}_N^α portfolio.

Let us now look at the \mathbf{w}_N^β portfolio. If the weights decrease at any rate slower than $1/N$, then the systematic exposure explodes because the factors are pervasive. On the other hand, if the weights decrease faster than $1/N$, then the portfolio risk declines to zero, leading to a Sharpe ratio of zero because the expression for the Sharpe ratio is exactly the same as the one for the risk of the portfolio. So, the rate of $1/N$ strikes the correct balance between optimizing the risk and return of the \mathbf{w}_N^β portfolio. Notice that the rate $1/N$ makes the \mathbf{w}_N^β portfolio well diversified, even with respect to idiosyncratic exposure, enhancing its Sharpe ratio even further.

We now present the main result of this section: under certain conditions, the beta portfolio can be replaced, without any loss of efficiency, by a class of *benchmark* portfolios that by construction are independent of the mean and variance of the observed factors, $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$, and hence, immune to beta misspecification.

Theorem 5 (Weight and Sharpe ratio of robust mean-variance portfolio for large N). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumptions 1 and 2 and $\boldsymbol{\alpha}_N \neq 0$. Suppose further that the investor holds a well-diversified benchmark portfolio $\mathbf{w}_N^{\text{bench}}$ satisfying the following properties:*

$$(\mathbf{w}_N^{\text{bench}})' \boldsymbol{\alpha}_N \rightarrow 0, \quad \mathbf{B}'_N \mathbf{w}_N^{\text{bench}} \rightarrow \mathbf{c}^{\text{bench}} \neq \mathbf{0}, \quad (\mathbf{w}_N^{\text{bench}})' \Sigma_N \mathbf{w}_N^{\text{bench}} \rightarrow 0, \quad (18)$$

where \mathbf{c}^{bench} is a $K \times 1$ vector of constants, different from the zero vector, satisfying $\boldsymbol{\lambda}'\mathbf{c}^{bench} \neq 0$.

(i) If \mathbf{c}^{bench} is perfectly proportional to the vector $\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$, then

$$\mathbf{w}_N^{rmv} \sim \frac{1}{\gamma\phi}\mathbf{w}_N^\alpha + \frac{1}{\gamma}\mathbf{w}_N^{bench},$$

without any loss of performance in terms of the Sharpe ratio from replacing \mathbf{w}_N^β with \mathbf{w}_N^{bench} .

(ii) If \mathbf{c}^{bench} is not proportional to the vector $\boldsymbol{\Omega}^{-1}\boldsymbol{\lambda}$, and \mathbf{w}_N^β and \mathbf{w}_N^{bench} have positive expected returns, then there is a loss of performance in terms of Sharpe ratio from replacing \mathbf{w}_N^β with \mathbf{w}_N^{bench} .

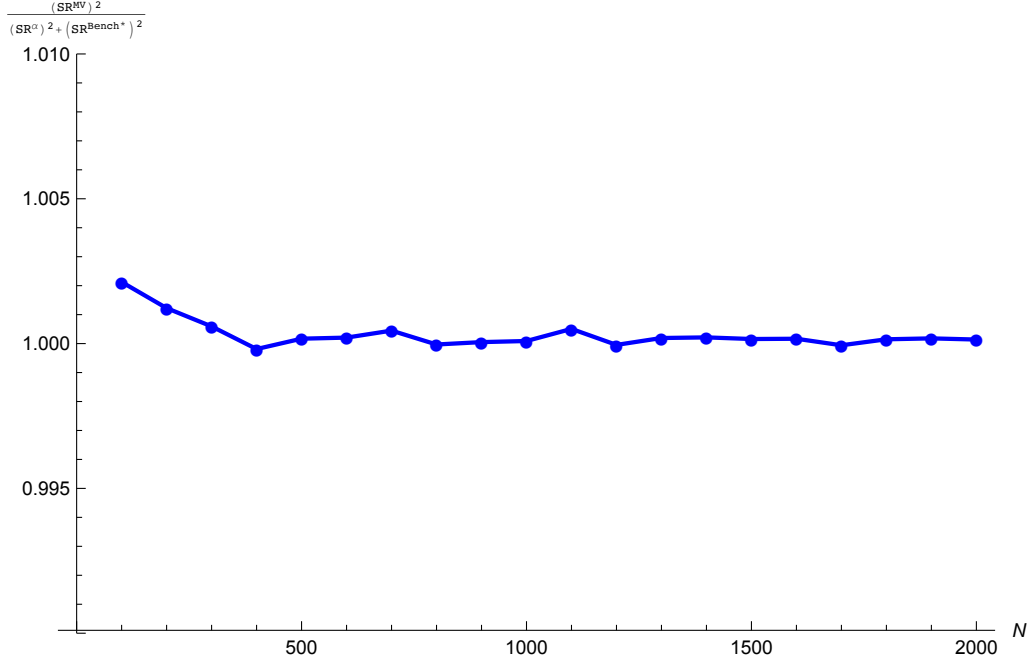
The first assumption in (18) implies that the benchmark portfolio is asymptotically orthogonal to $\boldsymbol{\alpha}_N$. The second assumption rules out that the benchmark portfolio return is equal to the risk-free return in the limit. The third assumption requires that the benchmark portfolio be well diversified. Note that for the case of large pricing errors (i.e., unbounded residual-variation case), the first assumption is satisfied whenever $(\mathbf{w}_N^{bench})'\mathbf{a}_N \rightarrow 0$ and $(\mathbf{w}_N^{bench})'\mathbf{A}_N \rightarrow 0$, where the latter condition ensures that \mathbf{w}_N^{bench} diversifies away the contribution of the latent factors, \mathbf{A}_N , to $\boldsymbol{\Sigma}_N$.

To construct a valid benchmark portfolio, one can rely on the insights of Treynor and Black (1973) and DeMiguel, Garlappi, and Uppal (2009). The results of Treynor and Black (1973) can be interpreted as saying that \mathbf{w}_N^β can be approximated by a portfolio that is similar to the market portfolio, \mathbf{w}_N^{mkt} . Our theorem formalizes the condition under which such an approximation will not lead to a loss in performance. This condition is always satisfied when there is only a single factor, that is, $K = 1$.

The assumptions in (18) imply that the return on the benchmark portfolio is asymptotically equivalent to the return on the portfolio of factors with weight \mathbf{c}^{bench} ; that is, $(\mathbf{w}_N^{bench})'(\mathbf{r}_t - r_f\mathbf{1}_N) = (\mathbf{c}^{bench})'(\mathbf{f}_t - r_f\mathbf{1}_K) + o_p(1)$. This choice guarantees that such a benchmark portfolio will achieve the same Sharpe ratio as the one that you would achieve if you knew the true risk premia, $\boldsymbol{\lambda}$, and the true covariance matrix for the factors, $\boldsymbol{\Omega}$.

Figure 4: Sharpe ratio of \mathbf{w}_N^α and $\mathbf{w}_N^{\text{bench}}$ relative to that of \mathbf{w}_N^{mv} portfolio

This figure shows that as the number of assets, N , increases, the sum of the squares of the Sharpe ratios of the \mathbf{w}_N^α and $\mathbf{w}_N^{\text{bench}}$ portfolios converges to the square of the Sharpe ratio of the mean-variance portfolio.



4.3 Mean-variance efficiency of robust portfolios

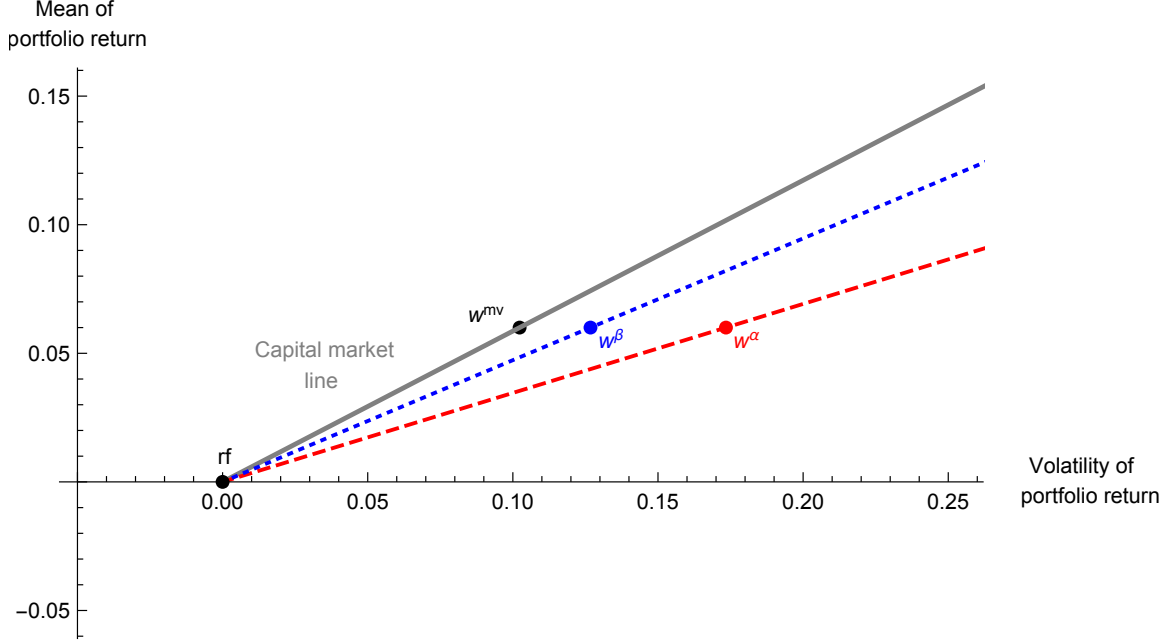
Above, we have established that the *robust* mean-variance portfolio obtained from the max-min optimization can be represented as a combination of two portfolios, \mathbf{w}_N^α and \mathbf{w}_N^β . Our portfolio addresses model misspecification by identifying a decision rule that works well *across* a set of models in the neighborhood of the approximating model. Of course, the robust portfolio is not necessarily optimal for *each* member of that set of models. Despite this, our robust portfolio, $\mathbf{w}_N^{\text{rmv}}$ is related to *mean-variance efficiency*, that is, the frontier of mean-variance returns achievable in the absence of concerns for misspecification.

In particular, we show in Theorem C3 that, under Assumptions 1 and 2, the traditional mean-variance portfolio weights (that is, the Markowitz weights obtained in the absence of concerns about misspecification) satisfy the following decomposition:

$$\mathbf{w}_N^{\text{mv}} = \frac{1}{\gamma} \mathbf{V}^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \sim \frac{1}{\gamma} \mathbf{w}_N^\alpha + \frac{1}{\gamma} \mathbf{w}_N^\beta,$$

Figure 5: Decomposition of the mean-variance portfolio

In this figure, we plot the mean-variance portfolio in the presence of risk-free asset, \mathbf{w}_N^{mv} , and its decomposition into two *inefficient* portfolios: one that depends only on the pricing errors, $\mathbf{w}_N^{\alpha_0}$, and another that depends only on the factor exposure and their premia, \mathbf{w}_N^{β} .



where the approximation relies on assuming that T is large enough. The above decomposition is not arbitrary: \mathbf{w}_N^{β} is the *minimum-variance portfolio* whose return is orthogonal to the return of \mathbf{w}_N^{α} and vice versa. More importantly, as shown in Theorem C3, two-fund separation holds: the portfolios \mathbf{w}_N^{α} and \mathbf{w}_N^{β} , which themselves are *inefficient*, can generate all the portfolios on the *efficient* mean-variance frontier of risky assets. The mean-variance portfolio and its decomposition into the “alpha” and “beta” portfolios, is displayed in Figure 5.

On the other hand, the *robust* mean-variance portfolio, $\mathbf{w}_N^{\text{rmv}}$, is a combination of \mathbf{w}_N^{α} and \mathbf{w}_N^{β} but will not be on the efficient frontier because from (10) we see that $\phi > 1$. Consequently, the weight on the \mathbf{w}_N^{α} portfolio in \mathbf{w}_N^{mv} is $1/\gamma$ while the weight on the \mathbf{w}_N^{β} portfolio in $\mathbf{w}_N^{\text{rmv}}$ is $1/(\phi\gamma)$. On the other hand, when considering the special case in (12), then $\phi \rightarrow 1$ as $T \rightarrow \infty$, and efficiency of the robust mean-variance portfolio will be achieved.

5 Simulation Experiment

In this section, our goal is to illustrate the *improvement* in out-of-sample portfolio performance that results from our new theoretical insights and to identify the circumstances when these improvements will be large and when they will be small. This section is divided into two parts. In the first part, we explain the design of our simulation experiment. In the second part, we demonstrate the improvement in out-of-sample performance resulting from using our robust portfolio model compared to using the equally weighted portfolio and the mean-variance Markowitz portfolio based on sample moments.

5.1 Simulation design and performance evaluation

The design of our simulation analysis is similar to that in MacKinlay and Pástor (2000). We consider the case in which the number of assets is $N = 100$; to illustrate the effect from having a large number of assets, we contrast the results for $N = 100$ with those for the case where $N = 5$. Throughout all our experiments, the investor assumes, and therefore estimates, a single-factor model:

$$\mathbf{r}_t = \boldsymbol{\alpha}_N + \boldsymbol{\beta}_N f_t + \boldsymbol{\varepsilon}_t.$$

We assume that the risk-free interest rate is 0 and that the observed factor f_t is IID and has Gaussian distribution. For the “base case” of our simulation exercise, we assume that the observed factor has a monthly mean equal to $\lambda = \frac{8}{12 \times 100}$ and monthly variance equal to $\Omega = (\frac{16}{\sqrt{12 \times 100}})^2$; both λ and Ω are scalars in our numerical illustration, because the investor assumes that there is only a single factor.

We measure the performance of a particular portfolio strategy using its out-of-sample Sharpe ratio, which is computed as follows. First, using the above parameter values, we simulate $M = 100$ paths of length $T = 300$ months. For each path, we estimate the parameters of the APT model using a rolling window of 120 monthly return observations, based on which we construct the portfolio weight for each of the portfolio strategies described below. We then compute the return of the portfolio strategy based on the realized return in the 121st month following the estimation window. We repeat the computation of the model parameters and portfolio weight for each of the subsequent 179 months, and then construct

the Sharpe ratio for this particular path of the simulation. We then undertake the same procedure for each simulated path and report the average Sharpe ratio across all the paths.

We consider the following portfolio strategies. (1) EW, the equal-weighted portfolio where the weights are equal to $1/N$ and no estimation is required; (2) MV, the mean-variance portfolio based on plugging in the sample mean and sample covariance matrix; (3) MLU, the ML-based *unconstrained* portfolio that is based on the sample mean and the spherical covariance matrix for the residuals: $\Sigma_N = \sigma_\varepsilon^2 \mathbf{I}_N$; (4) MLC, the ML-based *constrained* portfolio that imposes the APT restriction imposing that the investor uses the 99% quantile, $\chi_{N,99\%}^2$, in (15); (5) MLC_{bench}, just like MLC, but replacing the beta-portfolio with the benchmark $1/N$ portfolio; (6) TRUE, in which, instead of estimating the model, we use the population mean and covariance matrix to estimate the mean-variance portfolio; this portfolio allows us to assess the Sharpe ratio that would be achievable in the absence of estimation error.

We consider two environments, one with mispricing unrelated to common factors ($\alpha_N = \mathbf{a}_N$) and the other that has also one missing pervasive factor ($\alpha_N = \mathbf{A}_N \lambda_{miss} + \mathbf{a}_N$).²⁵ Both \mathbf{a}_N and \mathbf{A}_N are generated from an IID multivariate Gaussian distribution with mean $\mathbf{0}_N$ and covariance matrix equal to $\sigma_\alpha^2 \mathbf{I}_N$, with $\sigma_\alpha = \frac{1}{2} \frac{5}{\sqrt{12 \times 100}}$, which, in order to be conservative and consistent with the empirical data for individual stocks, is half of the value used in MacKinlay and Pástor (2000).²⁶ In both cases, ε_t is IID with a multivariate Gaussian distribution with a monthly mean of $\mathbf{0}_N$. In the case of small pricing errors unrelated to factors, the monthly covariance matrix of ε_t is $\Sigma_N = \sigma_\varepsilon^2 \mathbf{I}_N$; in contrast, for the case of large pricing errors related to pervasive factors, the monthly covariance matrix is $\Sigma_N = \mathbf{A}_N \mathbf{A}'_N + \sigma_\varepsilon^2 \mathbf{I}_N$. In both cases, $\sigma_\varepsilon = \frac{20}{\sqrt{12 \times 100}}$. Observe that to ensure identification, just as in MacKinlay and Pástor (2000), we set the variance of the missing factor equal to one, which implies that the risk premium λ_{miss} coincides with the Sharpe ratio for the missing factor; we set $\lambda_{miss} = \frac{0.75}{\sqrt{12}}$.

An important element of the estimation procedure is the choice of δ that appears in the APT constraint in equations (7) and (8). Ross (1976, p. 354) suggests constraining δ to be a multiple of the Sharpe ratio of the market portfolio. Instead, we propose the use of

²⁵The multi-step procedure outlined above in Section 3.3, based either on principal-component analysis or the approach in Gagliardini, Ossola, and Scaillet (2017), can be used to identify the presence of missing factors.

²⁶As we show below, using a larger value of σ_α would strengthen our results.

a 10-fold cross validation procedure in Hastie, Tibshirani, and Wainwright (2015, Section 2.3) to identify δ . Given that δ is not known, we identify it by maximizing the Sharpe ratio of the holdout portfolio returns in the cross-valuation approach.

5.2 Results

In this section, we describe the results for two cases of model misspecification that we study. We start by looking at Table 1, which studies the case of small pricing errors that are unrelated to factors and reports the annualized Sharpe ratios for the various portfolio strategies along with the t -statistic for the difference in the Sharpe ratio of the each portfolio relative to that of MLC_{bench} when there are $N = 100$ risky assets.

We start by looking at Panel A for the “base case” in Table 1. We see that the EW portfolio strategy achieves an annual Sharpe ratio of 0.54, which is higher than that of the MV portfolio. The poor performance of the MV portfolio has been well-documented in the literature (see DeMiguel, Garlappi, and Uppal (2009)) and the EW portfolio achieves a higher Sharpe ratio because it does not suffer from estimation error. Examining the MLU portfolio, we see that it performs much better than the EW portfolio, where the improvement is a result of essentially “shrinking” the residual-covariance matrix by setting $\Sigma_N = \sigma_\varepsilon^2 \mathbf{I}_N$, as described in MacKinlay and Pástor (2000). The MLC portfolio, which imposes the APT constraint, performs even better, which highlights the importance of imposing the APT restriction. Finally, the superior performance of the MLC_{bench} portfolio relative to MLC illustrates the additional gains that are possible from exploiting our insights about the beta portfolio when the number of risky assets is large. In order to highlight the asymptotic results, in Panel E we consider the case where $N = 5$. Comparing to the results in Panel A for $N = 100$, we find that when the number of risky assets is small, MLC_{bench} no longer outperforms MLC .

In addition to the “base case” of the simulations described above, we look at three variations in Panels B, C, and D of Table 1. Panel B considers the case in which the true risk premium on the observed factor, λ , is set to half its base-case value, which corresponds to a low-return environment. Even in this setting, the Sharpe ratio of MLC_{bench} continues to be significantly higher than that of MLU as well as that of EW and MV. Panel C considers the case in which the residual risk is half of its base-case value. In this case, the Sharpe ratio of the EW portfolio does not change at all, but the Sharpe ratios of the

Table 1: Out-of-Sample Sharpe Ratios with Model Misspecification

This table reports, for the case in which mispricing is unrelated to common factors ($\alpha_N = \mathbf{a}_N$), the annualized Sharpe ratios averaged across $M = 100$ Monte Carlo simulations for the following strategies: EW, the equal-weighted portfolio; MV, the mean-variance portfolio based on the sample mean and sample covariance matrix; MLU, the ML-based unconstrained mean-variance portfolio based on the sample mean and covariance matrix implied by the factor model, $\beta_N \beta_N' \Omega + \sigma_\epsilon^2 \mathbf{I}_N$ but without the APT constraint; MLC, the ML-based constrained mean-variance portfolio based on the sample mean and factor covariance matrix of $\beta_N \beta_N' \Omega + \Sigma_N$ with the APT constraint in which δ is obtained using ten-fold cross validation; MLC_{bench} , which is the same as MLC but with the benchmark portfolio replacing the beta portfolio. The t -statistic is for the difference in the Sharpe ratio of each strategy with respect to the Sharpe ratio of the MLC_{bench} portfolio.

	EW	MV	MLU	MLC	MLC_{bench}	True
<i>Panel A: Base case with $N = 100$</i>						
SR	0.54	0.40	1.08	1.69	2.00	3.94
t-stat wrt MLC_b	14.37	15.64	8.66	3.15		
<i>Panel B: Lower λ (half of base case)</i>						
SR	0.29	0.40	1.02	1.18	1.25	3.92
t-stat wrt MLC_b	9.93	9.09	2.09	0.80		
<i>Panel C: Lower σ_ϵ (half of base case)</i>						
SR	0.54	0.83	1.86	3.99	4.69	7.84
t-stat wrt MLC_b	20.83	18.60	11.67	4.70		
<i>Panel D: Lower σ_α (half of base case)</i>						
SR	0.51	0.17	0.49	0.75	0.72	2.01
t-stat wrt MLC_b	5.06	11.55	5.22	-0.73		
<i>Panel E: Base case with $N = 5$</i>						
SR	0.47	0.47	0.53	0.39	0.34	
t-stat wrt MLC_b	-3.31	-3.22	-4.66	-2.05		

other strategies, which rely on estimated return moments, improve substantially. Finally, one might wonder under what conditions the EW strategy will outperform the MLC_{bench} strategy: this happens when the alphas are small in absolute value. To illustrate this, in Panel D, we consider the case in which the dispersion of alphas is only half the base-case value. In this case, the Sharpe ratio of the EW strategy is still the same as its base-case value, but now the Sharpe ratio of MLC_{bench} is smaller, though still significantly greater than that of the EW and MLU portfolios.

We now examine the case of large pricing errors related to pervasive factors in Table 2. We see that the insights for even this case, with pricing errors that are related to factors,

Table 2: Out-of-Sample Sharpe Ratios with Beta Misspecification

This table reports, for the case in which mispricing has both a component unrelated to common factors and a component related to one missing pervasive factor ($\alpha_N = \mathbf{A}_N \lambda_{miss} + \mathbf{a}_N$), the annualized Sharpe ratios averaged across $M = 100$ Monte Carlo simulations for the following strategies: EW, the equal-weighted portfolio; MV, the mean-variance portfolio based on the sample mean and sample covariance matrix; MLU, the ML-based unconstrained mean-variance portfolio based on the sample mean and covariance matrix implied by the factor model, $\beta_N \beta_N' \Omega + \sigma_\epsilon^2 \mathbf{I}_N$ but without the APT constraint; MLC, the ML-based constrained mean-variance portfolio based on the sample mean and factor covariance matrix of $\beta_N \beta_N' \Omega + \Sigma_N$ with the APT constraint in which δ is obtained using ten-fold cross validation; MLC_{bench} , which is the same as MLC but with the benchmark portfolio replacing the beta portfolio. The t -statistic is for the difference in the Sharpe ratio of each strategy with respect to the Sharpe ratio of the MLC_{bench} portfolio.

	EW	MV	MLU	MLC	MLC_{bench}	True
<i>Panel A: Base case with $N = 100$</i>						
SR	0.46	0.43	1.10	1.62	1.52	4.22
t-stat wrt MLC_b	9.20	9.63	3.30	-0.96		
<i>Panel B: Lower λ (half of base case)</i>						
SR	0.21	0.42	1.14	1.21	1.24	4.20
t-stat wrt MLC_b	11.86	10.09	1.66	0.35		
<i>Panel C: Lower σ_ϵ (half of base case)</i>						
SR	0.48	0.81	3.49	3.63	4.11	8.36
t-stat wrt MLC_b	14.95	13.13	2.15	1.48		
<i>Panel D: Lower σ_α (half of base case)</i>						
SR	0.48	0.15	0.43	0.80	0.97	2.17
t-stat wrt MLC_b	8.96	15.68	7.79	5.82		
<i>Panel E: Base case with $N = 5$</i>						
SR	0.48	0.46	0.47	0.04	0.01	0.76
t-stat wrt MLC_b	-15.15	-12.23	-12.31	1.45		

are similar to the ones reported in Table 1. In particular, MLC and MLC_{bench} outperform the other strategies. However, the Sharpe ratios of the MLC and MLC_{bench} are similar. The reason for this is that the pricing errors are more precisely estimated thanks to the presence of the loading on the missing factor in both the expected return and the residual covariance matrix. This has two consequences. First, it means that the alpha portfolio contributes more to the overall performance relative to the beta portfolio, and therefore, the gains from improving the estimation of the beta portfolio are smaller. Second, it means that the beta portfolio itself is more precisely estimated because the covariance matrix of

returns is estimated much more precisely, implying that the gains from using the benchmark portfolio are smaller. The results in Panel E for the case of $N = 5$ shows this.

6 Conclusion

In this paper, we have provided a rigorous foundation and characterization, based on the APT, for alpha and beta portfolios, where the “alpha” portfolio is one that depends on pricing errors and the “beta” portfolio depends on factor risk premia. We then show how these properties can be exploited to mitigate the effects of model misspecification for portfolio choice.

Our first result is to explain that one can extend the interpretation of the APT so that the alpha in it represents not just small pricing errors that are independent of factors but also large pricing errors arising from mismeasured or missing factors. We also show how the APT model can capture misspecification in the beta component of returns. We then use the mathematical structure underlying the APT to study the mitigation of model misspecification for the mean-variance portfolio.

Our key insight is that instead of treating model misspecification directly in the mean-variance portfolios, it is better to first decompose mean-variance portfolios into two components, a “beta” portfolio and an “alpha” portfolio, and then to treat misspecification in these two components using different methods. Misspecification in the beta component of returns is treated by utilizing the property that the weights of the alpha portfolio dominate the corresponding weights in the beta portfolio as the number of assets increases asymptotically. Misspecification in the alpha component of returns is treated by imposing the APT restriction on the weighted sum of squares of the pricing errors when estimating the return-generating model. We use simulations to demonstrate that these theoretical findings can be exploited to achieve an improvement in out-of-sample portfolio performance that is both economically and statistically significant.

A Proofs for Theorems

Proof of Theorem 1

By Chamberlain and Rothschild (1983, Theorem 4) the residual covariance matrix satisfies

$$\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N,$$

where \mathbf{C}_N is a positive definite matrix with eigenvalues uniformly bounded by $g_{p+1N}(\boldsymbol{\Sigma}_N)$. By the Sherman-Morrison-Woodbury decomposition,

$$\boldsymbol{\Sigma}_N^{-1} = \mathbf{C}_N^{-1} - \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1}.$$

Therefore, by substitution,

$$\begin{aligned} \boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N &= \boldsymbol{\alpha}'_N \mathbf{C}_N^{-1} \boldsymbol{\alpha}_N - \boldsymbol{\alpha}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \boldsymbol{\alpha}_N \\ &= (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N)' \mathbf{C}_N^{-1} (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N) \\ &\quad - (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N)' \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} (\mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N) \\ &= \boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\ &\quad + \mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \\ &\quad + 2\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - 2\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss}. \end{aligned}$$

We now show that $\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N$ is bounded even as N diverges. We look each of the term on the right-hand side of the last equality sign, one by one. Thus,

$$\begin{aligned} &\boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \boldsymbol{\lambda}'_{miss} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\ &= \boldsymbol{\lambda}'_{miss} (\mathbf{I}_N - \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1}) \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\ &= \boldsymbol{\lambda}'_{miss} (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \leq \boldsymbol{\lambda}'_{miss} \boldsymbol{\lambda}_{miss}, \end{aligned}$$

because $\mathbf{I}_p - (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N$ is positive semidefinite. Next, for the third term,

$$\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \leq \mathbf{a}'_N \mathbf{a}_N g_{NN}^{-1}(\mathbf{C}_N).$$

Now, the j th element of $\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N$, obtained by considering the j th column of \mathbf{A}_N , for every $1 \leq j \leq p$, satisfies

$$|\mathbf{a}'_N \mathbf{C}_N^{-1} g_{jN}^{\frac{1}{2}} \mathbf{v}_{jN}| \leq g_{jN}^{\frac{1}{2}} (\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{a}_N)^{\frac{1}{2}} (\mathbf{v}'_{jN} \mathbf{C}_N^{-1} \mathbf{v}_{jN})^{\frac{1}{2}} \leq g_{jN}^{\frac{1}{2}} g_{NN}^{-1}(\mathbf{C}_N) (\mathbf{a}'_N \mathbf{a}_N)^{\frac{1}{2}},$$

recalling that $\mathbf{v}'_{jN} \mathbf{v}_{jN} = 1$, where for simplicity we set $\mathbf{v}_{jN} = \mathbf{v}_{jN}(\boldsymbol{\Sigma}_N)$, $g_{jN} = g_{jN}(\boldsymbol{\Sigma}_N)$.

Moreover, the (i, j) th element, for every $1 \leq i, j \leq p$, of $(\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ is equal to $g_{iN}^{\frac{1}{2}} g_{jN}^{\frac{1}{2}} \mathbf{v}'_{iN} \mathbf{C}_N^{-1} \mathbf{v}_{jN}$. Therefore, assuming without loss of generality that $g_{1N} = \max[g_{1N}, \dots, g_{pN}]$ for N large

enough, then $(\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1}$ decreases at rate g_{1N}^{-1} . On the other hand, for the same reason, the elements of the vector $\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N$ diverge at most at rate $g_{1N}^{\frac{1}{2}}$. Thus, the fourth term satisfies:

$$|\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N| \leq \delta g_{1N}^{\frac{1}{2}} g_{1N}^{\frac{1}{2}} g_{1N}^{-1} = \delta.$$

Concerning the last two terms, it turns out that their difference converges to zero. In fact,

$$\begin{aligned} & |2\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - 2\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss}| \\ &= 2|\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \boldsymbol{\lambda}_{miss}| \\ &\leq (\mathbf{a}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N)^{\frac{1}{2}} (\boldsymbol{\lambda}'_{miss} (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \boldsymbol{\lambda}_{miss})^{\frac{1}{2}} \\ &\leq \delta g_{pN}^{-\frac{1}{2}} \rightarrow 0. \end{aligned}$$

■

Proof of Theorem 2

Re-write the *max-min*-optimization as

$$\max_{\mathbf{w}_N} \min_{\boldsymbol{\alpha}_N} \left\{ \mathbf{w}'_N (\boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}) - \frac{\gamma}{2} \mathbf{w}_N \mathbf{V}_N \mathbf{w}_N \right\}.$$

We first show that the relative entropy constraint satisfies

$$\begin{aligned} & \int_{-\infty}^{\infty} \left(\ln \frac{f_{\boldsymbol{\alpha}}(\mathbf{r}_t^e)}{f_{\hat{\boldsymbol{\alpha}}}(\mathbf{r}_t^e)} \right) f_{\boldsymbol{\alpha}}(\mathbf{r}_t^e) d\mathbf{r}_t^e \\ &= \frac{1}{2} \left(\hat{\boldsymbol{\alpha}}'_N \mathbf{V}_N^{-1} \hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}'_N \mathbf{V}_N^{-1} \boldsymbol{\alpha}_N - 2\hat{\boldsymbol{\alpha}}'_N \mathbf{V}_N^{-1} E(\mathbf{r}_t^e - \mathbf{B}_N \boldsymbol{\lambda}) + 2\boldsymbol{\alpha}'_N \mathbf{V}_N^{-1} E(\mathbf{r}_t^e - \mathbf{B}_N \boldsymbol{\lambda}) \right) \\ &= \frac{1}{2} \left(\hat{\boldsymbol{\alpha}}'_N \mathbf{V}_N^{-1} \hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}'_N \mathbf{V}_N^{-1} \boldsymbol{\alpha}_N - 2\hat{\boldsymbol{\alpha}}'_N \mathbf{V}_N^{-1} \boldsymbol{\alpha}_N + 2\boldsymbol{\alpha}'_N \mathbf{V}_N^{-1} \boldsymbol{\alpha}_N \right) \\ &= \frac{1}{2} (\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N)' \mathbf{V}_N^{-1} (\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N) \leq \frac{1}{2} (\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N)' \boldsymbol{\Sigma}_N^{-1} (\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N) \leq \delta, \end{aligned}$$

where the first inequality arises from noticing that, by the Sherman decomposition, $\mathbf{V}_N^{-1} = \boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1}$. Considering first the minimization step, one gets the Lagrangian

$$\mathcal{L}(\boldsymbol{\alpha}_N, \epsilon) = \mathbf{w}'_N \boldsymbol{\alpha}_N + \frac{\epsilon}{2} [(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)' \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) - 2\delta],$$

where we need to find the saddle-point satisfying $\mathcal{L}(\boldsymbol{\alpha}_N^{rmv}, \epsilon^{rmv}) = \min_{\boldsymbol{\alpha}_N} \max_{\epsilon \geq 0} \mathcal{L}(\boldsymbol{\alpha}_N, \epsilon)$. Consider first the case when $\epsilon > 0$. Then, the first-order condition (with respect to $\boldsymbol{\alpha}_N$) that can be re-written as:

$$0 = \mathbf{w}_N + \epsilon \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\alpha}_N^{rmv} - \hat{\boldsymbol{\alpha}}_N),$$

and re-arranging gives

$$\boldsymbol{\alpha}_N^{\text{rmv}} = \hat{\boldsymbol{\alpha}}_N - \frac{\boldsymbol{\Sigma}_N}{\epsilon} \mathbf{w}_N.$$

Substituting the constraint $(\boldsymbol{\alpha}_N^{\text{rmv}} - \hat{\boldsymbol{\alpha}}_N)' \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\alpha}_N^{\text{rmv}} - \hat{\boldsymbol{\alpha}}_N) = 2\delta$ into $\mathbf{w}_N = \epsilon \boldsymbol{\Sigma}_N^{-1} (\hat{\boldsymbol{\alpha}}_N - \boldsymbol{\alpha}_N^{\text{rmv}})$, and re-arranging, yields the final solution

$$\boldsymbol{\alpha}_N^{\text{rmv}} = \hat{\boldsymbol{\alpha}}_N - \frac{(2\delta)^{\frac{1}{2}}}{(\mathbf{w}'_N \boldsymbol{\Sigma}_N \mathbf{w}_N)^{\frac{1}{2}}} \boldsymbol{\Sigma}_N \mathbf{w}_N, \quad \epsilon^{\text{rmv}} = \sqrt{\frac{\mathbf{w}'_N \boldsymbol{\Sigma}_N \mathbf{w}_N}{2\delta}}.$$

We now show that the Lagrangian is globally minimized at this point, ruling out the case $\epsilon = 0$. In fact,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}_N, 0) &= \mathbf{w}'_N \boldsymbol{\alpha}_N = \mathbf{w}'_N \hat{\boldsymbol{\alpha}}_N + \mathbf{w}'_N (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) \\ &\geq \mathbf{w}'_N \hat{\boldsymbol{\alpha}}_N - |\mathbf{w}'_N (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)| \geq \mathbf{w}'_N \hat{\boldsymbol{\alpha}}_N - (\mathbf{w}'_N \boldsymbol{\Sigma}_N \mathbf{w}_N)^{\frac{1}{2}} ((\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)' \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N))^{\frac{1}{2}} \\ &> \mathbf{w}'_N \hat{\boldsymbol{\alpha}}_N - (\mathbf{w}'_N \boldsymbol{\Sigma}_N \mathbf{w}_N)^{\frac{1}{2}} (2\delta)^{\frac{1}{2}} = \mathcal{L}(\boldsymbol{\alpha}_N^{\text{rmv}}, \epsilon^{\text{rmv}}), \end{aligned}$$

where the strict inequality follows by the slackness condition for $\epsilon = 0$. The last term on the right hand side is precisely the Lagrangian evaluated when the relative entropy constraint is binding, that is for $\epsilon = \epsilon^{\text{rmv}} > 0$.

Consider now the maximization step:

$$\begin{aligned} &\max_{\mathbf{w}_N} \left\{ \mathbf{w}'_N (\boldsymbol{\alpha}_N^{\text{rmv}} + \mathbf{B}_N \boldsymbol{\lambda}) - \frac{\gamma}{2} \mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N \right\} \\ &= \max_{\mathbf{w}_N} \left\{ \mathbf{w}'_N \left(\hat{\boldsymbol{\alpha}}_N - \frac{(2\delta)^{\frac{1}{2}}}{(\mathbf{w}'_N \boldsymbol{\Sigma}_N \mathbf{w}_N)^{\frac{1}{2}}} \boldsymbol{\Sigma}_N \mathbf{w}_N + \mathbf{B}_N \boldsymbol{\lambda} \right) - \frac{\gamma}{2} \mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N \right\} \\ &= \max_{\mathbf{w}} \left\{ \mathbf{w}'_N (\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda}) - (2\delta)^{\frac{1}{2}} (\mathbf{w}'_N \boldsymbol{\Sigma}_N \mathbf{w}_N)^{1/2} - \frac{\gamma}{2} \mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N \right\}. \end{aligned}$$

By re-arranging the first-order condition:

$$0 = \hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda} - (2\delta)^{\frac{1}{2}} (\mathbf{w}_N^{\text{rmv}' / \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}})^{-1/2} \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}} - \gamma \mathbf{V}_N \mathbf{w}_N^{\text{rmv}},$$

one obtains $\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda} = \left(\left(\frac{2\delta}{(\mathbf{w}_N^{\text{rmv}' / \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}})} \right)^{\frac{1}{2}} \boldsymbol{\Sigma}_N + \gamma \mathbf{V}_N \right) \mathbf{w}_N^{\text{rmv}}$. Recalling that $\mathbf{V}_N = \boldsymbol{\Sigma}_N + \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N$, one obtains

$$\begin{aligned} \mathbf{w}_N^{\text{rmv}} &= \left(\left[\gamma + \left(\frac{2\delta}{(\mathbf{w}_N^{\text{rmv}' / \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}})} \right)^{\frac{1}{2}} \right] \boldsymbol{\Sigma}_N + \gamma \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N \right)^{-1} (\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda}) \\ &= \frac{1}{\gamma} \left(\underbrace{\left[1 + \left(\frac{2\delta}{\gamma^2 (\mathbf{w}_N^{\text{rmv}' / \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}})} \right)^{\frac{1}{2}} \right]}_{=\phi} \boldsymbol{\Sigma}_N + \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N \right)^{-1} (\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda}) \end{aligned}$$

$$= \frac{1}{\gamma} (\phi \boldsymbol{\Sigma}_N + \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N)^{-1} (\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda}),$$

where

$$\phi = \left[1 + \left(\frac{2\delta}{\gamma^2 (\mathbf{w}_N^{\text{rmv}'} \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}})} \right)^{\frac{1}{2}} \right].$$

■

Setting, $\delta = \frac{1}{2} \frac{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{T} \chi_{N, x\%}^2$, then the entropy constraint can be easily re-written as

$$T(\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N)' \frac{\boldsymbol{\Sigma}_N^{-1}}{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})} (\boldsymbol{\alpha}_N - \hat{\boldsymbol{\alpha}}_N) \leq \chi_{N, x\%}^2 < \infty,$$

which defines the set of values for $\boldsymbol{\alpha}_N$ that are statistically indistinguishable from $\hat{\boldsymbol{\alpha}}_N$ with probability $x\%$ or, equivalently, the $x\%$ confidence set for the true $\boldsymbol{\alpha}_N$ (centered around $\hat{\boldsymbol{\alpha}}_N$). Note that $(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}) \boldsymbol{\Sigma}_N$ represents the asymptotic covariance matrix of the Gaussian MLE $\hat{\boldsymbol{\alpha}}_N$.

Corollary 1. *Under the assumptions of Theorem 2 and $\delta = \frac{1}{2} \frac{(1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{T} \chi_{N, x\%}^2$, where $\chi_{N, x\%}^2$ denotes the x^{th} ($0 \leq x \leq 1$) quantile of a χ_N^2 distribution the shrinkage parameters ϕ becomes equal to:*

$$\phi = \left[1 + \left(\frac{\chi_{N, x\%}^2 (1 + \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{\gamma^2 T (\mathbf{w}_N^{\text{rmv}'} \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}})} \right)^{\frac{1}{2}} \right].$$

Proof of Theorem 3

For part (i), by the Sherman-Morrison-Woodbury formula, one obtains

$$\begin{aligned} \mathbf{w}_N^{\text{rmv}} &= \frac{1}{\gamma \phi} \left(\boldsymbol{\Sigma}_N + \frac{1}{\phi} \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N \right)^{-1} (\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda}) \\ &= \frac{1}{\gamma \phi} \left(\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\phi \boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \right) (\hat{\boldsymbol{\alpha}}_N + \mathbf{B}_N \boldsymbol{\lambda}) \\ &= \frac{1}{\gamma \phi} \left(\boldsymbol{\Sigma}_N^{-1} \hat{\boldsymbol{\alpha}}_N - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\phi \boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \hat{\boldsymbol{\alpha}}_N \right) \\ &\quad + \frac{1}{\gamma \phi} \left(\boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\phi \boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N \right) \boldsymbol{\lambda} \\ &= \frac{1}{\gamma \phi} \left(\boldsymbol{\Sigma}_N^{-1} \hat{\boldsymbol{\alpha}}_N - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\phi \boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \hat{\boldsymbol{\alpha}}_N \right) \\ &\quad + \frac{1}{\gamma} \left(\boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\phi \boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \right) \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}. \end{aligned}$$

For large N one obtains (element by element)

$$\begin{aligned}\mathbf{w}_N^{\text{rmv}} &\sim \frac{1}{\gamma\phi} \left(\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \right) \hat{\boldsymbol{\alpha}}_N \\ &\quad + \frac{1}{\gamma} \left(\boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \right) \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} \\ &= \frac{1}{\gamma\phi} \boldsymbol{\Sigma}_N^+ \hat{\boldsymbol{\alpha}}_N + \frac{1}{\gamma} \mathbf{V}_N^{-1} \mathbf{B}_N \boldsymbol{\lambda}.\end{aligned}$$

Result (ii) follows from the fact the \mathbf{w}_N^β diversifies idiosyncratic risk, namely that $\mathbf{w}_N^\beta \boldsymbol{\Sigma}_N \mathbf{w}_N^\beta \rightarrow 0$, unlike \mathbf{w}_N^α , which instead only diversifies away common risk. Therefore, $\mathbf{w}_N^{\text{rmv}'} \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}} \sim (\gamma\phi)^{-2} \mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N^+ \mathbf{w}_N^\alpha$, yielding

$$\begin{aligned}\phi &= \left[1 + \left(\frac{2\delta}{\gamma^2 (\mathbf{w}_N^{\text{rmv}'} \boldsymbol{\Sigma}_N \mathbf{w}_N^{\text{rmv}})} \right)^{\frac{1}{2}} \right] \\ &\sim \left[1 + \left(\frac{2\delta}{\gamma^2 (\gamma\phi)^{-2} \mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N^+ \mathbf{w}_N^\alpha} \right)^{\frac{1}{2}} \right] = \left[1 + \left(\frac{(2\delta)^{\frac{1}{2}} |\phi|}{(\mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N^+ \mathbf{w}_N^\alpha)^{\frac{1}{2}}} \right) \right].\end{aligned}$$

This provides a nonlinear equation in ϕ (due to the modulus) but, by seeking only the positive solution $\phi > 0$, one obtains the (approximate) linear equation

$$\phi \sim \left[1 + \left(\frac{(2\delta)^{\frac{1}{2}} \phi}{(\mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N^+ \mathbf{w}_N^\alpha)^{\frac{1}{2}}} \right) \right],$$

with solution (ignoring the large- N approximation) $\phi = [1 - (2\delta)^{\frac{1}{2}} (\mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N^+ \mathbf{w}_N^\alpha)^{-\frac{1}{2}}]^{-1}$, assuming (without loss of generality) T large enough, where $(\mathbf{w}_N^{\alpha'} \boldsymbol{\Sigma}_N^+ \mathbf{w}_N^\alpha) = (\hat{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \hat{\boldsymbol{\alpha}}_N)$ is bounded by Lemma C4. ■

Proof of Theorem 4

(i) We now consider the case where there are missing pervasive factors; the case with only small pricing errors that are unrelated to factors follows by setting $\mathbf{A}_N = \mathbf{0}$. Assumption $\boldsymbol{\alpha}_N \neq \mathbf{0}$ rules out that $\mathbf{w}_N^\alpha = \mathbf{0}$ for any finite N . Recall that now $\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$ and $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}} + \mathbf{a}_N$. Assuming a large enough T so that $\hat{\boldsymbol{\alpha}}_N \sim \boldsymbol{\alpha}_N$, consider first \mathbf{w}_N^α where its i th component satisfies:

$$w_{N,i}^\alpha = \mathbf{1}'_{N_i} \mathbf{w}_N^\alpha \sim \mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^+ \boldsymbol{\alpha}_N = \mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N - \mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N,$$

where $\mathbf{1}_{N_i}$ is an N -dimensional vector in which the i th element is one and the rest of the elements are zero. We deal with the two terms on the right-hand side of $w_{N,i}^\alpha$ separately. By the

Sherman-Morrison-Woodbury formula $\Sigma_N^{-1} = \mathbf{C}_N^{-1} - \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1}$, obtaining

$$\begin{aligned}
\mathbf{1}'_{N_i} \Sigma_N^{-1} \boldsymbol{\alpha}_N &= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \boldsymbol{\alpha}_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \boldsymbol{\alpha}_N \\
&= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\
&\quad + \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \\
&= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \boldsymbol{\lambda}_{miss} \\
&\quad + \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N.
\end{aligned}$$

By Holder's inequality, taking the norm and using the relation between norm and maximum eigenvalue, one obtains

$$|\mathbf{1}'_{N_i} \Sigma_N^{-1} \boldsymbol{\alpha}_N| = \mathcal{O} \left(\|\boldsymbol{\lambda}_{miss}\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\|}{f(N)} + |\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N| + \|\mathbf{a}_N\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\|}{f^{\frac{1}{2}}(N)} \right).$$

Along the same lines

$$\begin{aligned}
\mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N &= \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N - \mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{B}_N, \\
\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N &= \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{B}_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{B}_N, \quad \text{and} \\
\mathbf{B}'_N \Sigma_N^{-1} \boldsymbol{\alpha}_N &= \mathbf{B}'_N \mathbf{C}_N^{-1} \hat{\boldsymbol{\alpha}}_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \boldsymbol{\alpha}_N \\
&= \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{miss} \\
&\quad + \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N \\
&= \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \boldsymbol{\lambda}_{miss} \\
&\quad + \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{a}_N - \mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{A}_N (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{a}_N.
\end{aligned}$$

Therefore, using the same arguments as above, one obtains

$$|\mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} \boldsymbol{\alpha}_N| = \mathcal{O} \left(\|\mathbf{a}_N\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\| + \|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)} \right),$$

because, under our assumptions, the eigenvalues of $\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N$ and $\mathbf{B}'_N \mathbf{C}_N^{-1} \mathbf{B}_N$ have the same behavior. In particular, the first term $\mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N$ is $\mathcal{O}(\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\| + \|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N\|)$, the second term $\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N$ is $\mathcal{O}(f(N))$, and the third term $\mathbf{B}'_N \Sigma_N^{-1} \boldsymbol{\alpha}_N$ is $\mathcal{O}(f^{\frac{1}{2}}(N) \|\mathbf{a}_N\|)$.

For the \mathbf{w}_N^β portfolio, its i th component satisfies

$$\begin{aligned}
w_{N,i}^\beta &= \frac{1}{\gamma^\beta} \mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N \boldsymbol{\lambda} - \frac{1}{\gamma^\beta} \mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N \boldsymbol{\lambda} \\
&= \frac{1}{\gamma^\beta} \mathbf{1}'_{N_i} \Sigma_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda},
\end{aligned}$$

and using the above formulae for $\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N$ and $\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N$ concludes, where we use $\boldsymbol{\lambda} \neq \mathbf{0}$.

Regarding part (ii), $\mathbf{w}'_N \boldsymbol{\alpha}_N \leq \|\boldsymbol{\Sigma}_N^+\| (\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^+ \boldsymbol{\alpha}_N) \leq \delta < \infty$. Moreover, $\boldsymbol{\lambda}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{V}_N^{-1} \mathbf{B}_N \boldsymbol{\lambda} = \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} = \mathcal{O}(1/f(N))$ because $\|\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\| \leq \|\boldsymbol{\Sigma}_N^{-1}\| \|\mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|$. This implies $|(\mathbf{w}'_N)^\beta \mathbf{w}_N^\beta| = \mathcal{O}(1/f(N))$, which means that the squared norm of \mathbf{w}_N^β goes to zero.

Part (iii) follows from $|\mathbf{1}'_N \mathbf{w}_N^\alpha| \leq (\mathbf{1}'_N \boldsymbol{\Sigma}_N^+ \mathbf{1}_N)^{\frac{1}{2}} (\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^+ \boldsymbol{\alpha}_N)^{\frac{1}{2}}$ with $\mathbf{1}'_N \boldsymbol{\Sigma}_N^+ \mathbf{1}_N \rightarrow \infty$. In fact $\mathbf{1}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N \geq N/(g_{1N}(\boldsymbol{\Sigma}_N)) \geq 0$. On the other hand, $|\mathbf{1}'_N \mathbf{w}_N^\beta| = \frac{1}{\gamma^\beta} |(\mathbf{1}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N) (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}| \leq \delta < \infty$. ■

For the case where the pricing error is small (unrelated to factors), the result in (17) follows from the fact that, under Assumptions 1 and 2, the absolute value of the components of the mean-variance portfolio vectors decrease at most at the rate:

$$|w_{N,i}^\alpha| = \mathcal{O} \left(|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N| + \frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)} \right), \quad (\text{A1})$$

$$|w_{N,i}^\beta| = \mathcal{O} \left(\frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f(N)} \right). \quad (\text{A2})$$

From equations (A1) and (A2), we see that $w_{N,i}^\alpha$ can dominate $w_{N,i}^\beta$ as the number of assets increases. In particular, $w_{N,i}^\alpha$ dominates $w_{N,i}^\beta$ when the pricing-error term, $|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N|$, goes to zero slowly as the number of assets increases. The weights $w_{N,i}^\alpha$ dominate $w_{N,i}^\beta$ also because the second term on the right-hand side of (A1), $\frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)}$, dominates the term on the right-hand side of (A2), $\frac{\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N\|}{f(N)}$; this dominance arises because the two terms have different denominators: $f^{\frac{1}{2}}(N)$ instead of $f(N)$.

Recall that the APT bounds the pricing error from above; that is, $\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N \leq \delta < \infty$. However, the APT is silent about whether $\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N$ is bounded away from zero. When this expression is bounded away from zero, one can show that the ratio $w_{N,i}^\beta/w_{N,i}^\alpha$ always decreases at a rate that is equal or faster than $1/f^{\frac{1}{2}}(N)$.

Note that for the case with pricing errors that are related to pervasive factors (that is, where $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N$ and $\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$), the result in (17) follows from the fact that, under Assumptions 1 and 2, the absolute value of the components of the

mean-variance portfolio vectors decrease at most at the rate

$$|w_{N,i}^\alpha| = \mathcal{O}\left(\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N\| + \|\mathbf{a}_N\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\| + \|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N\|}{f^{\frac{1}{2}}(N)} + \|\boldsymbol{\lambda}_{miss}\| \frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\|}{f(N)}\right); \quad (\text{A3})$$

$$|w_{N,i}^\beta| = \mathcal{O}\left(\frac{\|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{A}_N\| + \|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{B}_N\|}{f(N)}\right). \quad (\text{A4})$$

Recall that $\boldsymbol{\lambda}_{miss}$ can be interpreted as the risk premia on the unobserved factors with loadings \mathbf{A}_N , whereas the vector \mathbf{a}_N represents the pure pricing error that is not associated with a factor structure. The \mathbf{a}_N component dominates the behavior of the portfolio weights $w_{N,i}^\alpha$ in (A3), whereas the risk premia $\boldsymbol{\lambda}_{miss}$ component declines to zero faster. In general, the portfolio weight $w_{N,i}^\beta$ in (A4) declines at the same, fast rate as the risk premia $\boldsymbol{\lambda}_{miss}$ component of $w_{N,i}^\alpha$. When \mathbf{a}_N is non-zero then, as before, the \mathbf{w}_N^α portfolio dominates the \mathbf{w}_N^β portfolio across all three norms considered in the theorem above.

Observe that Equation (17) shows that the portfolio \mathbf{w}_N^α dominates \mathbf{w}_N^β for large N . Observe that \mathbf{w}_N^α is functionally independent of the factor risk premia, $\boldsymbol{\lambda}$, and the factor covariance matrix, $\boldsymbol{\Omega}$, making it robust to misspecification in the beta component of returns by construction. In contrast, portfolio \mathbf{w}_N^β depends on both $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$.

In order to get a better understanding of the result in (17), in the corollary below we look at a special case where $f(N) = N$. We consider only part (i) of the theorem, because the other parts of the theorem are unchanged under the special case.

Corollary 2 (Weights of alpha and beta portfolios for large N : Special case). *Suppose that the assumptions of Theorem 4 are satisfied and that the row sums of \mathbf{A}_N , \mathbf{B}_N and \mathbf{C}_N^{-1} are uniformly bounded.²⁷ Suppose also that $f(N) = N$. Then, as $N \rightarrow \infty$:*

- *for the case of small pricing errors that are unrelated to factors, the absolute value of the components of the mean-variance portfolio vectors \mathbf{w}_N^α and \mathbf{w}_N^β decrease at most at the rate²⁸*

$$|w_{N,i}^\alpha| = \mathcal{O}\left(\|\mathbf{1}'_{N_i} \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N\| + \frac{1}{N^{\frac{1}{2}}}\right) \quad \text{and} \quad |w_{N,i}^\beta| = \mathcal{O}\left(\frac{1}{N}\right), \quad \text{and}$$

- *for the case of pricing errors related to pervasive factors, the absolute value of the components of the mean-variance portfolio vectors \mathbf{w}_N^α and \mathbf{w}_N^β decrease at most at*

²⁷Given an $N \times M$ matrix \mathbf{D} , we say its row sums are uniformly bounded when $\sup_{1 \leq j \leq N} \sum_{i=1}^M |d_{ji}| \leq \delta < \infty$, for some arbitrary δ .

²⁸For any finite dimensional non-negative a_N and b_N , $a_N = \mathcal{O}(b_N)$ means that $a_N/b_N \leq \delta < \infty$, for some constant $\delta > 0$.

the rate

$$|w_{N,i}^\alpha| = \mathcal{O}\left(|\mathbf{1}'_{N_i} \mathbf{C}_N^{-1} \mathbf{a}_N| + \frac{\|\mathbf{a}_N\|}{N^{\frac{1}{2}}} + \frac{\|\boldsymbol{\lambda}_{miss}\|}{N}\right) \quad \text{and} \quad |w_{N,i}^\beta| = \mathcal{O}\left(\frac{1}{N}\right).$$

Proof of Theorem 5

The mean and variance of the excess return for the benchmark portfolio satisfy $\mathbf{w}_N^{bench'}(\boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}) \rightarrow \mathbf{c}^{bench'} \boldsymbol{\lambda}$ and $\mathbf{w}_N^{bench'} \mathbf{V}_N \mathbf{w}_N^{bench} \rightarrow \mathbf{c}^{bench'} \boldsymbol{\Omega} \mathbf{c}^{bench}$, respectively.

Consider first the case when $\mathbf{c}^{bench} = \delta \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$, for some scalar $\delta \neq 0$. Then

$$\lim_{N \rightarrow \infty} (SR^{bench})^2 \rightarrow (SR_\infty^{bench})^2 = \frac{\delta^2 (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})^2}{\delta^2 (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})} = \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda} = (SR_\infty^\beta)^2.$$

In fact, by easy calculations, the beta portfolio satisfies

$$\mu^\beta - r_f = \boldsymbol{\lambda}' \mathbf{B}'_N \mathbf{V}_N^{-1} (\boldsymbol{\alpha}_N + \mathbf{B}_N \boldsymbol{\lambda}), \quad \sigma^\beta = (\boldsymbol{\lambda}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \boldsymbol{\lambda})^{\frac{1}{2}},$$

yielding $SR^\beta \rightarrow (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})^{\frac{1}{2}} = SR_\infty^\beta$, because $\boldsymbol{\lambda}' \mathbf{B}'_N \mathbf{V}_N^{-1} \boldsymbol{\alpha}_N \rightarrow 0$ and $\boldsymbol{\lambda}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$. Note that, by Lemma C1, $(SR_\infty^{rmv})^2 \leq (SR_\infty^\alpha)^2 + (SR_\infty^{bench})^2$, where SR_∞^α is the limit (as $N \rightarrow \infty$) of $\hat{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^+ \boldsymbol{\alpha}_N / (\hat{\boldsymbol{\alpha}}'_N \boldsymbol{\Sigma}_N^+ \hat{\boldsymbol{\alpha}}_N)^{\frac{1}{2}}$, bounded by Lemma C4.

Now consider the case when \mathbf{c}^{bench} is not proportional to $\boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$. Then

$$(SR_\infty^{bench})^2 = \frac{((\mathbf{c}^{bench})' \boldsymbol{\lambda})^2}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} < (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}) = (SR_\infty^\beta)^2,$$

because

$$\frac{((\mathbf{c}^{bench})' \boldsymbol{\lambda})^2}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} = \frac{((\mathbf{c}^{bench})' \boldsymbol{\Omega}^{\frac{1}{2}} \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\lambda})^2}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} < \frac{((\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench})(\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})}{(\mathbf{c}^{bench})' \boldsymbol{\Omega} \mathbf{c}^{bench}} = (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}).$$

The strict inequality is implied whenever $\boldsymbol{\Omega}^{\frac{1}{2}} \mathbf{c}^{bench}$ and $\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\lambda}$ are not proportional, which in turn is equivalent to \mathbf{c}^{bench} being not being proportional to $\boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}$, as stated above. Now, by Lemma C1, the Sharpe ratio SR^* associated with the portfolio $\mathbf{w}_N^* = \mathbf{w}_N^\alpha / (\gamma \phi) + \mathbf{w}_N^{bench} / \gamma$ satisfies

$$(SR^*)^2 = (SR^\alpha)^2 + (SR^{bench})^2 - \frac{1}{\mathbf{w}_N^{*'} \mathbf{V}_N \mathbf{w}_N^*} (\sigma^{bench} SR^\alpha - \sigma^\alpha SR^{bench})^2 < (SR^{rmv})^2,$$

whenever

$$SR > -\frac{\sigma^\alpha}{\sigma^{bench}} SR^\alpha \quad \text{for every } SR^{bench} \leq SR \leq SR^\beta, \quad (\text{A5})$$

assuming N large enough such that $SR^\beta > 0$ (recall that $S_\infty^\beta = (\boldsymbol{\lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda})^{\frac{1}{2}}$). When the inequality in the left hand side of (A5) is satisfied, the first-derivative of $(SR^*)^2$ with respect

to SR^{bench} is positive, implying that $(SR^*)^2$ increases to $(SR^{rmv})^2$ when replacing SR^{bench} with SR^β . ■

B Estimation of approximating model

B.1 Estimation for the case of small pricing errors

In this section, we consider the case of small pricing errors that are unrelated to pervasive factors ($p = 0$). Our argument applies to virtually any (parametric) estimation procedure, but we will illustrate it with respect to the (pseudo) Gaussian ML estimator. This is a natural estimator for our model when the first two moments of asset returns are specified correctly, although distributional assumptions (such as normality) are not required; hence, the use of pseudo ML. The (pseudo) ML estimator, based on the unconditional joint distribution of $\begin{pmatrix} \mathbf{r}_t \\ \mathbf{f}_t \end{pmatrix} - r_f \mathbf{1}_{N+K}$ and assuming i.i.d. residuals for simplicity, is²⁹

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) &= -\frac{1}{2} \log(\det(\tilde{\boldsymbol{\Sigma}}_N)) \\ &\quad - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{r}_t - r_f \mathbf{1}_N - \tilde{\boldsymbol{\alpha}}_N - \tilde{\mathbf{B}}_N(\mathbf{f}_t - r_f \mathbf{1}_K) \right)' \tilde{\boldsymbol{\Sigma}}_N^{-1} \left(\mathbf{r}_t - r_f \mathbf{1}_N - \tilde{\boldsymbol{\alpha}}_N - \tilde{\mathbf{B}}_N(\mathbf{f}_t - r_f \mathbf{1}_K) \right) \\ &\quad - \frac{1}{2} \log(\det(\tilde{\boldsymbol{\Omega}})) - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{f}_t - r_f \mathbf{1}_K - \tilde{\boldsymbol{\lambda}} \right)' \tilde{\boldsymbol{\Omega}}^{-1} \left(\mathbf{f}_t - r_f \mathbf{1}_K - \tilde{\boldsymbol{\lambda}} \right), \end{aligned} \quad (\text{B1})$$

where $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\alpha}}_N', \text{vec}(\tilde{\mathbf{B}}_N)', \text{vech}(\tilde{\boldsymbol{\Sigma}}_N)', \tilde{\boldsymbol{\lambda}}', \text{vech}(\tilde{\boldsymbol{\Omega}})')$.³⁰ Therefore, the ML estimators for $\boldsymbol{\alpha}_N$, \mathbf{B}_N , $\boldsymbol{\Sigma}_N$ coincide with the OLS estimators, conditional on the realization of the factors. On the other hand, the ML estimators for $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$ are the sample mean and sample covariance of the factors \mathbf{f}_t .

However, because the APT restriction is not guaranteed to hold, one should consider the maximum-likelihood estimator subject to this restriction. Moreover, with the parameter $\boldsymbol{\alpha}_N$ constrained by the APT restriction, imposing such a constraint may lead to a more precise estimator of the true parameter values compared to the *unconstrained* estimator, which we label $\hat{\boldsymbol{\theta}}_{MLU}$. The ML-*constrained* estimator, $\hat{\boldsymbol{\theta}}_{MLC}$, is given below.

²⁹Notice that we have expressed the joint distribution as the product of a conditional distribution and a marginal distribution. Relaxing the i.i.d. assumption requires specification of time-varying conditional means, conditional variances, and conditional covariances.

³⁰Note that $\det(\cdot)$ denotes the determinant, $\text{vec}(\cdot)$ denotes the operator that stacks the columns of a matrix into a single column vector, and $\text{vech}(\cdot)$ denotes the operator that stacks the unique elements of the columns of a symmetric matrix into a single column vector.

Theorem B1 (Parameter estimation by imposing asset-pricing restriction: Case for small pricing error). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumption 1. Then:*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} L(\tilde{\boldsymbol{\theta}}) \quad \text{subject to} \quad \tilde{\boldsymbol{\alpha}}'_N \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\boldsymbol{\alpha}}_N \leq \delta,$$

where $L(\tilde{\boldsymbol{\theta}})$ is defined in (B1). If $\left(\sum_{t=1}^T \dot{\mathbf{f}}_t \dot{\mathbf{f}}_t'\right)$ is nonsingular, then $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\boldsymbol{\alpha}}'_{N,MLC}, \operatorname{vec}(\hat{\mathbf{B}}_{N,MLC})', \operatorname{vech}(\hat{\boldsymbol{\Sigma}}_{N,MLC})', \hat{\boldsymbol{\lambda}}'_{MLC}, \operatorname{vech}(\hat{\boldsymbol{\Omega}}_{MLC})')'$ exists, where:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{N,MLC} &= \frac{1}{1 + \hat{\kappa}} \left[\bar{\mathbf{r}} - r_f \mathbf{1}_N - \hat{\mathbf{B}}_{N,MLC} (\bar{\mathbf{f}} - r_f \mathbf{1}_K) \right], \\ \hat{\mathbf{B}}_{N,MLC} &= \left(\sum_{t=1}^T \dot{\mathbf{r}}_t \dot{\mathbf{f}}_t' \right) \left(\sum_{t=1}^T \dot{\mathbf{f}}_t \dot{\mathbf{f}}_t' \right)^{-1}, \quad \text{and} \\ \hat{\boldsymbol{\Sigma}}_{N,MLC} &= \frac{1}{T} \sum_{t=1}^T (\dot{\mathbf{r}}_t - \hat{\mathbf{B}}_{N,MLC} \dot{\mathbf{f}}_t) (\dot{\mathbf{r}}_t - \hat{\mathbf{B}}_{N,MLC} \dot{\mathbf{f}}_t)', \end{aligned}$$

in which $\hat{\kappa} \geq 0$ is the optimal value of the Karush-Kuhn-Tucker multiplier, $\bar{\mathbf{f}} = T^{-1} \sum_{t=1}^T \mathbf{f}_t$, $\bar{\mathbf{r}}_N = T^{-1} \sum_{t=1}^T \mathbf{r}_t$, $\dot{\mathbf{f}}_t = \mathbf{f}_t - r_f \mathbf{1}_K - \frac{1}{(1+\hat{\kappa})} (\bar{\mathbf{f}} - r_f \mathbf{1}_K)$, $\dot{\mathbf{r}}_t = \mathbf{r}_t - r_f \mathbf{1}_N - \frac{1}{(1+\hat{\kappa})} (\bar{\mathbf{r}}_N - r_f \mathbf{1}_N)$, and the MLC estimators $\hat{\boldsymbol{\lambda}}_{MLC}$ and $\operatorname{vech}(\hat{\boldsymbol{\Omega}}_{MLC})$ coincide with the sample mean and covariance of the factors \mathbf{f}_t .

Proof. The formulae for $\hat{\boldsymbol{\alpha}}_{N,MLC}$, $\hat{\mathbf{B}}_{N,MLC}$ and $\hat{\boldsymbol{\Sigma}}_{N,MLC}$ follow from solving the first-order conditions associated with the Lagrangian problem:

$$\{\hat{\boldsymbol{\theta}}_{MLC}, \hat{\kappa}\} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \underset{\tilde{\kappa} \geq 0}{\operatorname{argmax}} L(\tilde{\boldsymbol{\theta}}) - \tilde{\kappa} (\tilde{\boldsymbol{\alpha}}'_N \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\boldsymbol{\alpha}}_N - \delta).$$

Start with $\tilde{\kappa} = 0$. Then the MLC estimator for $\boldsymbol{\theta}$ coincides with the OLS estimator $\hat{\boldsymbol{\theta}}$, readily obtained by setting $\hat{\kappa} = 0$ in the formulae for $\hat{\boldsymbol{\theta}}_{MLC}$, and one needs to evaluate whether $\hat{\boldsymbol{\alpha}}'_N \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\alpha}}_N > \delta$. If the latter inequality holds, $\tilde{\kappa} = 0$ violates the complementary slackness condition and can be ruled out. Alternatively, when $\hat{\boldsymbol{\alpha}}'_N \hat{\boldsymbol{\Sigma}}_N^{-1} \hat{\boldsymbol{\alpha}}_N \leq \delta$, then we evaluate $L(\hat{\boldsymbol{\theta}})$ and move on to case $\tilde{\kappa} > 0$. In particular, by solving the first-order equation for $\hat{\boldsymbol{\alpha}}_{N,MLC}$ and $\hat{\kappa}$ sequentially, one gets:

$$(1 + \hat{\kappa})^2 = \frac{\left[\bar{\mathbf{r}} - r_f \mathbf{1}_N - \hat{\mathbf{B}}_{N,MLC} (\bar{\mathbf{f}} - r_f \mathbf{1}_K) \right]' \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \left[\bar{\mathbf{r}} - r_f \mathbf{1}_N - \hat{\mathbf{B}}_{N,MLC} (\bar{\mathbf{f}} - r_f \mathbf{1}_K) \right]}{\delta},$$

and now case $\tilde{\kappa} > 0$ is feasible only when the right hand side of the above equation is bigger than one. When this occurs, one then evaluates $L(\hat{\boldsymbol{\theta}}_{MLC})$. Note that now $\hat{\boldsymbol{\alpha}}_{N,MLC}$ will satisfy the constraint exactly, that is $\hat{\boldsymbol{\alpha}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\boldsymbol{\alpha}}_{N,MLC} = \delta$, by construction. Incidentally, note that the parameters' estimates $\hat{\mathbf{B}}_{N,MLC}$ and $\hat{\boldsymbol{\Sigma}}_{N,MLC}$ are function of $\hat{\kappa}$

so a (simple) iterative procedure is required. Alternatively, when the right hand of the above equation side is smaller than one, case $\tilde{\kappa} > 0$ is ruled out, by the complementary slackness condition, and one retains the OLS estimator $\hat{\boldsymbol{\theta}}$. Finally, when both cases $\tilde{\kappa} = 0$ and $\tilde{\kappa} > 0$ are feasible, one simply needs to compare $L(\hat{\boldsymbol{\theta}})$ with $L(\hat{\boldsymbol{\theta}}_{\text{MLC}})$, and select the estimate (either $\hat{\boldsymbol{\theta}}_{\text{MLC}}$ or $\hat{\boldsymbol{\theta}}$) that maximizes the log-likelihood function. However, in most cases, either $\tilde{\kappa} = 0$ or $\tilde{\kappa} > 0$ is feasible, and not the other, simplifying the solution of the Kuch Tucker problem.

Finally, for $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ and $\hat{\boldsymbol{\Omega}}_{\text{MLC}}$, one obtains precisely the sample mean and sample covariance matrix of \mathbf{f}_t . ■

The constrained estimator $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$ turns out to be precisely the ridge estimator for $\boldsymbol{\alpha}_N$, unless $\hat{\kappa} = 0$, in which case the OLS estimator is re-obtained. Besides $\hat{\boldsymbol{\alpha}}_{N,\text{MLC}}$, the estimators of $\hat{\mathbf{B}}_{N,\text{MLC}}$ and $\hat{\boldsymbol{\Sigma}}_{N,\text{MLC}}$ are also functions of $\hat{\kappa}$ because of the APT constraint, in contrast to $\hat{\boldsymbol{\lambda}}_{\text{MLC}}$ and $\hat{\boldsymbol{\Omega}}_{\text{MLC}}$, which are simply the sample mean and sample covariance of the \mathbf{f}_t because the APT constraint does not affect the distribution of the traded factors \mathbf{f}_t .

B.2 Estimation for the case of missing pervasive factors

The second case of alpha misspecification is when there are $p > 0$ missing pervasive factors. For the case in which the pricing error is unrelated to the missing factors, \mathbf{a}_N is zero in (6), we get that $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}}$ and $\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$, where $\boldsymbol{\lambda}_{\text{miss}}$ is the risk premia corresponding to the missing factors, and \mathbf{C}_N is an $N \times N$ positive-definite matrix with bounded eigenvalues that represents the covariance matrix of the pure idiosyncratic component of the error returns. Observe that $\boldsymbol{\alpha}_N$ is a component of the expected return, $\boldsymbol{\mu}_N$; likewise, $\boldsymbol{\Sigma}_N$ is a component of the return-covariance matrix, \mathbf{V}_N . Hence, \mathbf{A}_N appears in both the mean and covariance matrix of returns. MacKinlay and Pástor (2000) use this insight to improve the precision of the estimated \mathbf{A}_N parameters, which, in turn, substantially improves the performance of the estimated portfolio.³¹ Importantly, using the Sherman-Morrison-Woodbury formula, it follows that

$$\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N = \boldsymbol{\lambda}'_{\text{miss}} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N \boldsymbol{\lambda}_{\text{miss}} = \boldsymbol{\lambda}'_{\text{miss}} (\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N) \boldsymbol{\lambda}_{\text{miss}}.$$

³¹Note that because we are interpreting the missing factors as unobserved, without loss of generality, one can assume that $\mathbf{A}_N \mathbf{A}'_N$ represents the contribution of the missing factors to the residual variance $\boldsymbol{\Sigma}_N$ because the missing factors are assumed to be uncorrelated and have unit variance, leaving the risk premia $\boldsymbol{\lambda}_{\text{miss}}$ as free parameters to be estimated. MacKinlay and Pástor (2000) consider a different identification assumption. For $p = 1$, they estimate $\boldsymbol{\alpha}_N$ without distinguishing between \mathbf{A}_N and $\boldsymbol{\lambda}_{\text{miss}}$, implying that the contribution of the single missing factor to the return variance equals $\boldsymbol{\alpha}_N \boldsymbol{\alpha}'_N / (\text{SR}^{\text{miss}})^2$, where SR^{miss} is the Sharpe ratio of the missing factor.

Thus, $\boldsymbol{\alpha}'_N \boldsymbol{\Sigma}_N^{-1} \boldsymbol{\alpha}_N$ converges to $\boldsymbol{\lambda}'_{miss} \boldsymbol{\lambda}_{miss}$ as $N \rightarrow \infty$ because $(\mathbf{I}_p + \mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)^{-1} (\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ converges to the identity matrix, given that the missing factors are pervasive, implying that $(\mathbf{A}'_N \mathbf{C}_N^{-1} \mathbf{A}_N)$ is increasing without bound. This means that the APT restriction is always satisfied for the case of only missing pervasive factors (that is, the case in which $\mathbf{a}_N = \mathbf{0}$), once we recognize that $\boldsymbol{\Sigma}_N$ contains the loadings of the missing factors, \mathbf{A}_N .

However, for the general unbounded-variation case where the pricing error consists of *both* missing factors and a component that is unrelated to factors, $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N$, the APT restriction is not automatically satisfied. Therefore, when estimating the model, we need to impose the restriction: $\mathbf{a}'_N \mathbf{a}_N \leq \delta < \infty$ for any N , or equivalently, impose the restriction in (5). Under the same assumptions as above concerning the K observed factors \mathbf{f}_t , in particular that these represent traded asset returns, the true unconditional means and covariances of returns now satisfy the equations below, where \mathbf{C}_N has bounded maximum eigenvalue, in contrast to $\boldsymbol{\Sigma}_N$:

$$E(\mathbf{r}_t - r_f \mathbf{1}_N) = \boldsymbol{\mu}_N - r_f \mathbf{1}_N = \mathbf{a}_N + \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{B}_N \boldsymbol{\lambda}, \quad \text{var}(\mathbf{r}_t) = \mathbf{V}_N = \mathbf{B}_N \boldsymbol{\Omega} \mathbf{B}'_N + \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N.$$

As in the previous case of bounded variation, the joint log-likelihood function can be decomposed as follows:

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) = & -\frac{1}{2} \log(\det(\tilde{\mathbf{A}}_N \tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)) - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{r}_t - r_f \mathbf{1}_N - \tilde{\mathbf{A}}_N \tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_N (\mathbf{f}_t - r_f \mathbf{1}_K) \right)' \\ & \times (\tilde{\mathbf{A}}_N \tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)^{-1} \left(\mathbf{r}_t - r_f \mathbf{1}_N - \tilde{\mathbf{A}}_N \tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_N (\mathbf{f}_t - r_f \mathbf{1}_K) \right) \\ & - \frac{1}{2} \log(\det(\tilde{\boldsymbol{\Omega}})) - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{f}_t - r_f \mathbf{1}_K - \tilde{\boldsymbol{\lambda}} \right)' \tilde{\boldsymbol{\Omega}}^{-1} \left(\mathbf{f}_t - r_f \mathbf{1}_K - \tilde{\boldsymbol{\lambda}} \right). \end{aligned} \quad (\text{B2})$$

Without loss of generality, one can assume that the missing factors are uncorrelated with each other and have unit variance, achieving identification of \mathbf{A}_N . However, $\boldsymbol{\lambda}_{miss}$ and $\boldsymbol{\alpha}_N$ *cannot* be identified separately unless the APT restriction is imposed, as shown in Theorem B2 below.

Theorem B2 (Parameter estimation by imposing asset-pricing restriction: Case for large pricing error with traded factors). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumption 1 and that $\boldsymbol{\Sigma}_{f^e f^e} - \bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'}$ is nonsingular, where $\boldsymbol{\Sigma}_{f^e f^e} = T^{-1} \sum_{t=1}^T \mathbf{f}_t^e \mathbf{f}_t^{e'}$, $\bar{\mathbf{f}}^e = (\bar{\mathbf{f}}^e - r_f \mathbf{1}_K)$ with $\mathbf{f}_t^e = (\mathbf{f}_t - r_f \mathbf{1}_K)$. Then*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\tilde{\boldsymbol{\theta}}}{\text{argmax}} L(\tilde{\boldsymbol{\theta}}) \quad \text{subject to} \quad \tilde{\mathbf{a}}'_N \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\mathbf{a}}_N \leq \delta,$$

where $L(\tilde{\boldsymbol{\theta}})$ is defined in (B2), and $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\mathbf{a}}'_{N,MLC}, \hat{\boldsymbol{\lambda}}_{miss,MLC}, \text{vec}(\hat{\mathbf{A}}_{N,MLC})', \text{vec}(\hat{\mathbf{B}}_{N,MLC})', \text{vech}(\hat{\mathbf{C}}_{N,MLC})', \hat{\boldsymbol{\lambda}}'_{MLC}, \text{vech}(\hat{\boldsymbol{\Omega}}_{MLC})')'$.

(i) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} > 0$, then

$$\begin{aligned} \text{vec}(\hat{\mathbf{B}}_{N,MLC}) &= \left((\boldsymbol{\Sigma}_{f^e f^e} \otimes \mathbf{I}) - (\bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \otimes (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N)) \right)^{-1} \text{vec} \left(\boldsymbol{\Sigma}_{r^e f^e} - (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N) \bar{\mathbf{r}}_N^e \bar{\mathbf{f}}^{e'} \right), \\ \hat{\boldsymbol{\lambda}}_{miss,MLC} &= (\hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{A}}_{N,MLC})^{-1} \hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} (\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e), \\ \hat{\mathbf{a}}_{N,MLC} &= \frac{1}{\hat{\kappa} + 1} \left(\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e - \hat{\mathbf{A}}_{N,MLC} \hat{\boldsymbol{\lambda}}_{miss,MLC} \right), \end{aligned} \quad (\text{B3})$$

where $\hat{\boldsymbol{\Sigma}}_{N,MLC} = \hat{\mathbf{A}}_{N,MLC} \hat{\mathbf{A}}'_{N,MLC} + \hat{\mathbf{C}}_{N,MLC}$, $\boldsymbol{\Sigma}_{r^e f^e} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t^e \mathbf{f}_t^{e'}$, $\boldsymbol{\Sigma}_{f^e f^e} = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t^e \mathbf{f}_t^{e'}$, setting $\mathbf{r}_t^e = (\mathbf{r}_t - r_f \mathbf{1}_N)$, $\bar{\mathbf{r}}_N^e = (\bar{\mathbf{r}}_N - r_f \mathbf{1}_N)$. and

$$\mathbf{G}_N = \frac{1}{(\hat{\kappa} + 1)} \mathbf{I}_N + \frac{\hat{\kappa}}{(\hat{\kappa} + 1)} \hat{\mathbf{A}}_{N,MLC} (\hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{A}}_{N,MLC})^{-1} \hat{\mathbf{A}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1}.$$

Note that $\hat{\mathbf{A}}_{N,MLC}$ and $\hat{\mathbf{C}}_{N,MLC}$ do not admit a closed-form solution and, as before, $\hat{\boldsymbol{\lambda}}_{MLC}$ and $\hat{\boldsymbol{\Omega}}_{MLC}$ coincide with the sample mean and sample covariance of the factors \mathbf{f}_t .

(ii) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} = 0$ one can estimate only $\boldsymbol{\alpha}_N = \mathbf{A}_N \boldsymbol{\lambda}_{miss} + \mathbf{a}_N$ but not the two components separately, and one obtains

$$\hat{\boldsymbol{\alpha}}_{N,MLC} = \bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e,$$

and the expression for $\text{vec}(\hat{\mathbf{B}}_{N,MLC})$ can be obtained by setting $\hat{\kappa} = 0$ in (B3). The expressions for $\hat{\boldsymbol{\lambda}}_{MLC}$ and $\hat{\boldsymbol{\Omega}}_{MLC}$ are unchanged, and, as before, the expressions for the estimators of $\hat{\mathbf{A}}_{N,MLC}$ and $\hat{\mathbf{C}}_{N,MLC}$ do not admit a closed-form solution.

Proof. Within this proof, for simplicity, we do not use the $\tilde{\cdot}$ notation to denote feasible parameter values.

As explained in the proof to Theorem B1, defining by $\hat{\boldsymbol{\theta}}$ the MLC corresponding to $\tilde{\kappa} = 0$, this is unfeasible whenever $\hat{\mathbf{a}}'_N \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{a}}_N > \delta$. Similarly, case $\tilde{\kappa} > 0$ is unfeasible whenever, for every $\tilde{\kappa} > 0$,

$$\left(\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e - \hat{\mathbf{A}}_{N,MLC} \hat{\boldsymbol{\lambda}}_{miss,MLC} \right)' \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \left(\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e - \hat{\mathbf{A}}_{N,MLC} \hat{\boldsymbol{\lambda}}_{miss,MLC} \right) < \delta,$$

$$\text{because } (1 + \tilde{\kappa})^2 = \frac{\left[\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e - \hat{\mathbf{A}}_{N,MLC} \hat{\boldsymbol{\lambda}}_{miss,MLC} \right]' \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \left[\bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,MLC} \bar{\mathbf{f}}^e - \hat{\mathbf{A}}_{N,MLC} \hat{\boldsymbol{\lambda}}_{miss,MLC} \right]}{\delta}.$$

When both cases are feasible, the optimal value for the Karush-Kuhn-Tucker multiplier $\tilde{\kappa}$ will be greater, or equal to zero, depending on which case maximizes the log-likelihood, namely depending on whether $L(\hat{\boldsymbol{\theta}}_{MLC})$ or $L(\hat{\boldsymbol{\theta}})$ is largest, respectively. Note that when $\tilde{\kappa} > 0$ then $\hat{\mathbf{a}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{a}}_{N,MLC} = \delta$ by construction.

We now derive the formulae for the estimators. Assume for now that case $\hat{\kappa} > 0$ holds. Differentiating the penalized log-likelihood with respect to $\boldsymbol{\lambda}_{miss}$, \mathbf{a}_N , and the Lagrange

multiplier κ , the first $K + N$ equations (after some algebra) are:

$$\begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix} (\bar{\mathbf{r}}_N^e - \mathbf{B}_N \bar{\mathbf{f}}^e) = \begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N & \mathbf{A}'_{N,\text{MLC}} \boldsymbol{\Sigma}_{N,\text{MLC}}^{-1} \\ \mathbf{A}_N & (1 + \hat{\kappa}) \mathbf{I}_N \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\lambda}}_{\text{miss},\text{MLC}} \\ \hat{\mathbf{a}}_{N,\text{MLC}} \end{pmatrix},$$

where recall that $\boldsymbol{\Sigma}_N = \mathbf{A}_N \mathbf{A}'_N + \mathbf{C}_N$. It is straightforward to see that, because of the APT restriction, $\boldsymbol{\lambda}_{\text{miss}}$ and \mathbf{a}_N can now be identified separately, as long as $\hat{\kappa} > 0$. In fact, the above system of linear equations can be solved because the matrix pre-multiplying $\hat{\boldsymbol{\lambda}}_{\text{miss},\text{MLC}}$ and $\hat{\mathbf{a}}_{N,\text{MLC}}$ is non-singular for every $\hat{\kappa} > 0$, leading to the closed-form solution:

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_{\text{miss},\text{MLC}} &= (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} (\bar{\mathbf{r}}_N^e - \mathbf{B}_N \bar{\mathbf{f}}^e), \\ \hat{\mathbf{a}}_{N,\text{MLC}} &= \frac{1}{\hat{\kappa} + 1} (\bar{\mathbf{r}}_N^e - \mathbf{B}_N \bar{\mathbf{f}}^e - \mathbf{A}_N \hat{\boldsymbol{\lambda}}_{\text{miss},\text{MLC}}). \end{aligned}$$

Turning now to the first-order condition with respect to the generic (a, b) th element of \mathbf{B}_N , denoted by B_{ab} , one obtains,

$$-\frac{1}{T} \sum_{t=1}^T \mathbf{g}'_t \boldsymbol{\Sigma}_N^{-1} \left(-\frac{\partial \mathbf{B}_N}{\partial B_{ab}} \mathbf{f}_t^e + \mathbf{G}_N \frac{\partial \mathbf{B}_N}{\partial B_{ab}} \bar{\mathbf{f}}^e \right) = 0, \text{ with } 1 \leq a \leq N, 1 \leq b \leq K,$$

setting, for simplicity,

$$\mathbf{g}_t = \left(\mathbf{r}_t^e - \mathbf{G}_N \bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{N,\text{MLC}} \mathbf{f}_t^e + \mathbf{G}_N \hat{\mathbf{B}}_{N,\text{MLC}} \bar{\mathbf{f}}^e \right) \quad \text{and} \quad \mathbf{g} = \frac{1}{T} \sum_{t=1}^T \mathbf{g}_t,$$

with

$$\mathbf{G}_N = \frac{1}{(\hat{\kappa} + 1)} \mathbf{I}_N + \frac{\hat{\kappa}}{(\hat{\kappa} + 1)} \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1}.$$

Taking the vec operator for both sides of the first-order condition above gives

$$\frac{1}{T} \sum_{t=1}^T (\mathbf{f}_t^{e'} \otimes \mathbf{g}'_t \boldsymbol{\Sigma}_N^{-1}) \text{vec} \left(\frac{\partial \mathbf{B}_N}{\partial B_{ab}} \right) = (\bar{\mathbf{f}}^{e'} \otimes \mathbf{g}' \boldsymbol{\Sigma}_N^{-1} \mathbf{G}_N) \text{vec} \left(\frac{\partial \mathbf{B}_N}{\partial B_{ab}} \right), \text{ with } 1 \leq a \leq N, 1 \leq b \leq K,$$

which can be rewritten more succinctly as

$$\frac{1}{T} \sum_{t=1}^T \mathbf{f}_t^e \mathbf{g}'_t = \bar{\mathbf{f}}^e \mathbf{g}' \boldsymbol{\Sigma}_N^{-1} \mathbf{G}_N \boldsymbol{\Sigma}_N.$$

Next, recalling that $\boldsymbol{\Sigma}_{r^e f^e} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t^e \mathbf{f}_t^{e'}$, and $\boldsymbol{\Sigma}_{f^e f^e} = \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t^e \mathbf{f}_t^{e'}$, with $\boldsymbol{\Sigma}_{f^e r^e} = \boldsymbol{\Sigma}'_{r^e f^e}$, one obtains $\boldsymbol{\Sigma}_N^{-1} \mathbf{G}_N \boldsymbol{\Sigma}_N = \frac{1}{(\hat{\kappa} + 1)} \mathbf{I}_N + \frac{\hat{\kappa}}{(\hat{\kappa} + 1)} \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N = \mathbf{G}'_N$ and rearranging the above first order-condition gives

$$\boldsymbol{\Sigma}_{f^e r^e} - \bar{\mathbf{f}}^e \bar{\mathbf{r}}_N^{e'} \mathbf{G}'_N - \boldsymbol{\Sigma}_{f^e f^e} \hat{\mathbf{B}}'_{N,\text{MLC}} + \bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \hat{\mathbf{B}}'_{N,\text{MLC}} \mathbf{G}'_N - (\bar{\mathbf{f}}^e \bar{\mathbf{r}}_N^{e'} - \bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \mathbf{B}'_{N,\text{MLC}}) (\mathbf{I}_N - \mathbf{G}'_N) \mathbf{G}'_N =$$

$$\boldsymbol{\Sigma}_{f^e r^e} - \bar{\mathbf{f}}^e \bar{\mathbf{r}}^{e'} (2\mathbf{G}'_N - \mathbf{G}'_N \mathbf{G}'_N) - \boldsymbol{\Sigma}_{f^e f^e} \hat{\mathbf{B}}'_{N,MLC} + \bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \hat{\mathbf{B}}'_{N,MLC} (2\mathbf{G}'_N - \mathbf{G}'_N \mathbf{G}'_N) = \mathbf{0}.$$

Transposing both sides, taking the vec, and solving for $\text{vec}(\hat{\mathbf{B}}_{N,MLC})$ gives

$$\text{vec}(\hat{\mathbf{B}}_{N,MLC}) = \left((\boldsymbol{\Sigma}_{f^e f^e} \otimes \mathbf{I}_N) - (\bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \otimes (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N)) \right)^{-1} \text{vec} \left(\boldsymbol{\Sigma}_{r^e f^e} - (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N) \bar{\mathbf{r}}_N^e \bar{\mathbf{f}}^{e'} \right).$$

We need to show that a solution for $\hat{\mathbf{B}}_{N,MLC}$ exists. This requires one to establish that the matrix $\left((\boldsymbol{\Sigma}_{f^e f^e} \otimes \mathbf{I}_N) - (\bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \otimes (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N)) \right)$ is invertible. This matrix can be written as

$$\left((\boldsymbol{\Sigma}_{f^e f^e} \otimes \mathbf{I}_N) - (\bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \otimes (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N)) \right) = \left(((\boldsymbol{\Sigma}_{f^e f^e} - \bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'}) \otimes \mathbf{I}_N) + (\bar{\mathbf{f}}^e \bar{\mathbf{f}}^{e'} \otimes (\mathbf{I}_N - (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N))) \right).$$

The first matrix on the right hand side is non-singular. One then just needs to show that the second matrix is positive semi-definitive. This follows because, $\mathbf{I}_N - (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N) = (\mathbf{I}_N - \mathbf{G}_N)(\mathbf{I}_N - \mathbf{G}_N)$, and we show below that $(\mathbf{I}_N - \mathbf{G}_N)$ is positive semi-definite.

$$\begin{aligned} \mathbf{I}_N - \mathbf{G}_N &= \mathbf{I}_N - \frac{1}{(\hat{\kappa} + 1)} \mathbf{I}_N - \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}} \right) \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ &= \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}} \right) (\mathbf{I}_N - \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1}) \\ &= \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}} \right) \boldsymbol{\Sigma}_N (\boldsymbol{\Sigma}_N^{-1} - \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1}) \\ &= \left(\frac{\hat{\kappa}}{1 + \hat{\kappa}} \right) \boldsymbol{\Sigma}_N \boldsymbol{\Sigma}_N^{-1/2} (\mathbf{I}_N - \boldsymbol{\Sigma}_N^{-1/2} \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1/2}) \boldsymbol{\Sigma}_N^{-1/2}. \end{aligned}$$

The right-hand side is the product of positive-definite matrices, namely $\boldsymbol{\Sigma}_N$ and $\boldsymbol{\Sigma}_N^{-1/2}$, and of the matrix $\mathbf{I}_N - \boldsymbol{\Sigma}_N^{-1/2} \mathbf{A}_N (\mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N)^{-1} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1/2}$, which is a projection matrix orthogonal to $\boldsymbol{\Sigma}_N^{-1/2} \mathbf{A}_N$, and therefore, positive semidefinite.

Therefore, plugging $\hat{\mathbf{B}}_{N,MLC}$ into $\hat{\boldsymbol{\lambda}}_{miss,MLC}$ and $\hat{\mathbf{a}}_{N,MLC}$, one obtains that

$$\hat{\boldsymbol{\lambda}}_{miss,MLC} = \hat{\boldsymbol{\lambda}}_{miss}(\mathbf{A}_N, \mathbf{C}_N), \quad \hat{\mathbf{a}}_{N,MLC} = \hat{\mathbf{a}}_N(\mathbf{A}_N, \mathbf{C}_N) \quad \text{and} \quad \hat{\kappa} = \hat{\kappa}(\mathbf{A}_N, \mathbf{C}_N).$$

Substituting them, together with $\hat{\mathbf{B}}_{N,MLC}$, into $L(\boldsymbol{\theta}) - \kappa(\mathbf{a}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{a}_N - \delta)$, gives the concentrated likelihood function, which is a function of only \mathbf{A}_N and \mathbf{C}_N which will be maximized numerically, providing $\mathbf{A}_{N,MLC}$ and $\mathbf{C}_{N,MLC}$.

Suppose now that $\hat{\kappa} = 0$ holds, and recall that in this case the MLC is indicated by $\hat{\boldsymbol{\theta}}$. One can clearly obtain a unique solution for $(\mathbf{A}_N, \mathbf{I}_N) \begin{pmatrix} \hat{\boldsymbol{\lambda}}_{miss} \\ \hat{\mathbf{a}}_N \end{pmatrix} = \mathbf{A}_N \hat{\boldsymbol{\lambda}}_{miss} + \hat{\mathbf{a}}_N$.

However, to solve for $\hat{\boldsymbol{\lambda}}_{miss}$ and $\hat{\mathbf{a}}_N$ separately, one needs to invert the matrix

$$\begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ \mathbf{I}_N \end{pmatrix} (\mathbf{A}_N, \mathbf{I}_N) = \begin{pmatrix} \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{A}_N & \mathbf{A}'_N \boldsymbol{\Sigma}_N^{-1} \\ \mathbf{A}_N & \mathbf{I}_N \end{pmatrix},$$

which is not possible because it is of dimension $(N + K) \times (N + K)$ but of rank N , as the left-hand side shows that it is obtained from the product of two matrices of dimension $(N + K) \times N$. All the other parameters are identified separately, and their expressions follow from differentiating $L(\boldsymbol{\theta})$ and solving the resulting first-order conditions. For instance, the formula for $\hat{\mathbf{B}}_N$ follow by setting $\mathbf{G}_N = \mathbf{I}_N$ into (B3).

■

B.3 Estimation for the general case with missing pervasive factors and non-traded observed factors

We now present the general case when some of the observed factors are non-traded, together with the features discussed in the previous two sections, namely pricing errors unrelated to factors and missing pervasive factors. It turns out that the above results can be immediately generalized. In particular, assume now that

$$\mathbf{r}_t^e = \boldsymbol{\alpha}_N + \mathbf{B}_{1N}(\boldsymbol{\lambda}_1 + \mathbf{f}_{1t} - E(f_{1t})) + \mathbf{B}_{2N}\mathbf{f}_{2t}^e + \boldsymbol{\varepsilon}_t,$$

with

$$\boldsymbol{\alpha}_N = \mathbf{a}_N + \mathbf{A}_N\boldsymbol{\lambda}_{miss},$$

$$\text{var}(\mathbf{r}_t^e) = \mathbf{V}_N = \mathbf{B}_N\boldsymbol{\Omega}\mathbf{B}'_N + \mathbf{A}_N\mathbf{A}'_N + \mathbf{C}_N,$$

where we set $\mathbf{B}_N = (\mathbf{B}_{1N}, \mathbf{B}_{2N})$, $\mathbf{f}_t = (\mathbf{f}'_{1t}, \mathbf{f}'_{2t})'$, $\boldsymbol{\Omega} = \text{var}(\mathbf{f}_t)$ with \mathbf{f}_{1t} denoting the set of K_1 non-traded observed factors and \mathbf{f}_{2t}^e the set of K_2 traded observed factors, expressed as excess returns, where $K = K_1 + K_2$. Given that \mathbf{f}_{2t}^e are excess returns on traded assets, their risk premia satisfy $\boldsymbol{\lambda}_2 = E(\mathbf{f}_{2t}^e)$ and, to avoid confusion with the risk premia of the non-traded assets, we will use the expectation formulation for $\boldsymbol{\lambda}_2$.

The following theorem, reported without proof, extends Theorems B1 and B2, where the joint log-likelihood function takes now the form:

$$\begin{aligned} L(\tilde{\boldsymbol{\theta}}) = & -\frac{1}{2} \log(\det(\tilde{\mathbf{A}}_N\tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)) - \frac{1}{2T} \sum_{t=1}^T \left(\mathbf{r}_t^e - \tilde{\mathbf{A}}_N\tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_{1N}(\tilde{\boldsymbol{\lambda}}_1 + \mathbf{f}_{1t} - \widetilde{E(\mathbf{f}_{1t})}) - \tilde{\mathbf{B}}_{2N}(\mathbf{f}_{2t}^e) \right)' \\ & \times (\tilde{\mathbf{A}}_N\tilde{\mathbf{A}}'_N + \tilde{\mathbf{C}}_N)^{-1} \left(\mathbf{r}_t^e - \tilde{\mathbf{A}}_N\tilde{\boldsymbol{\lambda}}_{miss} - \tilde{\mathbf{a}}_N - \tilde{\mathbf{B}}_{1N}(\tilde{\boldsymbol{\lambda}}_1 + \mathbf{f}_{1t} - \widetilde{E(\mathbf{f}_{1t})}) - \tilde{\mathbf{B}}_{2N}(\mathbf{f}_{2t}^e) \right) \\ & - \frac{1}{2} \log(\det(\tilde{\boldsymbol{\Omega}})) - \frac{1}{2T} \sum_{t=1}^T \left((\mathbf{f}'_{1t}, \mathbf{f}'_{2t}) - (\widetilde{E(\mathbf{f}_{1t})}', \widetilde{E(\mathbf{f}_{2t}^e)'}) \right) \tilde{\boldsymbol{\Omega}}^{-1} \left((\mathbf{f}'_{1t}, \mathbf{f}'_{2t}) - (\widetilde{E(\mathbf{f}_{1t})}', \widetilde{E(\mathbf{f}_{2t}^e)'}) \right)'. \end{aligned} \tag{B4}$$

Without loss of generality, one can assume that the missing factors are uncorrelated with each other, and with the observed factors, and have unit variance to achieve identification of

\mathbf{A}_N . However, $\boldsymbol{\lambda}_{miss}$, $\boldsymbol{\lambda}_1$ and \mathbf{a}_N cannot be identified separately unless the APT restriction is imposed, generalizing our finding in Theorem B2.

Theorem B3 (Parameter estimation by imposing asset-pricing restriction: the general case). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumption 1 and that $\boldsymbol{\Sigma}_{f_2^e f_2^e} - \bar{\mathbf{f}}_2^e \bar{\mathbf{f}}_2^{e'}$ is nonsingular, where $\boldsymbol{\Sigma}_{f_2^e f_2^e} = T^{-1} \sum_{t=1}^T \mathbf{f}_{2t}^e \mathbf{f}_{2t}^{e'}$ and $\bar{\mathbf{f}}_2^e = T^{-1} \sum_{t=1}^T \mathbf{f}_{2t}^e$. Then*

$$\hat{\boldsymbol{\theta}}_{MLC} = \underset{\tilde{\boldsymbol{\theta}}}{\operatorname{argmax}} L(\tilde{\boldsymbol{\theta}}) \quad \text{subject to} \quad \tilde{\mathbf{a}}_N' \tilde{\boldsymbol{\Sigma}}_N^{-1} \tilde{\mathbf{a}}_N \leq \delta,$$

where $L(\tilde{\boldsymbol{\theta}})$ is defined in (B4), and $\hat{\boldsymbol{\theta}}_{MLC} = (\hat{\boldsymbol{\alpha}}'_{N,MLC}, \hat{\boldsymbol{\lambda}}'_{miss,MLC}, \hat{\boldsymbol{\lambda}}'_{1,MLC}, \widehat{E(\mathbf{f}_{1t})}_{MLC}, \widehat{E(\mathbf{f}_{2t}^e)}'_{MLC}, \operatorname{vec}(\hat{\mathbf{A}}_{N,MLC})', \operatorname{vec}(\hat{\mathbf{B}}_{N,MLC})', \operatorname{vech}(\hat{\mathbf{C}}_{N,MLC})', \operatorname{vech}(\hat{\boldsymbol{\Omega}}_{MLC})')'$.

(i) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} > 0$, setting

$$\mathbf{D}_N = (\mathbf{A}_N, \mathbf{B}_{1N}), \quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}'_{miss}, \boldsymbol{\lambda}'_1)',$$

then

$$\operatorname{vec}(\hat{\mathbf{B}}_{2N,MLC}) = \left((\boldsymbol{\Sigma}_{f_2^e f_2^e} \otimes \mathbf{I}) - (\bar{\mathbf{f}}_2^e \bar{\mathbf{f}}_2^{e'} \otimes (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N)) \right)^{-1} \operatorname{vec} \left(\boldsymbol{\Sigma}_{h, f_2^e} - (2\mathbf{G}_N - \mathbf{G}_N \mathbf{G}_N) \bar{\mathbf{h}}_N \bar{\mathbf{f}}_2^{e'} \right),$$

$$\hat{\boldsymbol{\lambda}}_{MLC} = (\hat{\mathbf{D}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{D}}_{N,MLC})^{-1} \hat{\mathbf{D}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} (\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,MLC} \bar{\mathbf{f}}_2^e), \quad (\text{B5})$$

$$\hat{\mathbf{a}}_{N,MLC} = \frac{1}{\hat{\kappa} + 1} \left(\bar{\mathbf{h}}_N - \hat{\mathbf{B}}_{2N,MLC} \bar{\mathbf{f}}_2^e - \hat{\mathbf{D}}_{N,MLC} \hat{\boldsymbol{\lambda}}_{MLC} \right),$$

where $\hat{\boldsymbol{\Sigma}}_{N,MLC} = \hat{\mathbf{A}}_{N,MLC} \hat{\mathbf{A}}'_{N,MLC} + \hat{\mathbf{C}}_{N,MLC}$, $\boldsymbol{\Sigma}_{h, f_2^e} = \frac{1}{T} \sum_{t=1}^T \mathbf{h}_t \mathbf{f}_{2t}^{e'}$, $\bar{\mathbf{h}}_N = \frac{1}{T} \sum_{t=1}^T \mathbf{h}_t$ with $\mathbf{h}_t = \mathbf{r}_t^e - \hat{\mathbf{B}}_{1N,MLC}(\mathbf{f}_{1t} - \bar{\mathbf{f}}_{1t})$, and

$$\mathbf{G}_N = \frac{1}{(\hat{\kappa} + 1)} \mathbf{I}_N + \frac{\hat{\kappa}}{(\hat{\kappa} + 1)} \hat{\mathbf{D}}_{N,MLC} (\hat{\mathbf{D}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1} \hat{\mathbf{D}}_{N,MLC})^{-1} \hat{\mathbf{D}}'_{N,MLC} \hat{\boldsymbol{\Sigma}}_{N,MLC}^{-1}.$$

Note that $\hat{\mathbf{D}}_{N,MLC} = (\hat{\mathbf{A}}_{N,MLC}, \hat{\mathbf{B}}_{1N,MLC})$ and $\hat{\mathbf{C}}_{N,MLC}$ do not admit a closed-form solution and, as before, $\widehat{E(\mathbf{f}_t)}_{MLC}$ and $\hat{\boldsymbol{\Omega}}_{MLC}$ coincide with the sample mean and sample covariance of the observed factors $\mathbf{f}_t = (\mathbf{f}'_{1t}, \mathbf{f}'_{2t})'$.

(ii) If the optimal value of the Karush-Kuhn-Tucker multiplier satisfies $\hat{\kappa} = 0$ one can estimate only $\boldsymbol{\alpha}_N = \mathbf{a}_N + \mathbf{D}_N \boldsymbol{\lambda}$ but not the three components separately, and one obtains

$$\hat{\boldsymbol{\alpha}}_{N,MLC} = \bar{\mathbf{r}}_N^e - \hat{\mathbf{B}}_{2N,MLC} \bar{\mathbf{f}}_2^e,$$

and the expression for $\operatorname{vec}(\hat{\mathbf{B}}_{2N,MLC})$ can be obtained by setting $\hat{\kappa} = 0$ in the terms that appear in (B5). The expressions for $\widehat{E(\mathbf{f}_t)}_{MLC}$ and $\hat{\boldsymbol{\Omega}}_{MLC}$ are unchanged, and, as before, the expressions for the estimators of $\hat{\mathbf{D}}_{N,MLC}$ and $\hat{\mathbf{C}}_{N,MLC}$ do not admit a closed-form solution.

C Auxiliary results

C.1 Decomposition of the Sharpe ratio

Theorem C1. Consider the portfolio weights $\mathbf{w}_N = \mathbf{w}_{N,1} + \mathbf{w}_{N,2}$ such that $\mathbf{w}_{N,1}$ is orthogonal to $\mathbf{w}_{N,2}$:

$$\mathbf{w}'_{N,1} \mathbf{V}_N \mathbf{w}_{N,2} = 0.$$

Then, defining $SR_i = \mathbf{w}'_{N,i}(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) / (\mathbf{w}'_{N,i} \mathbf{V}_N \mathbf{w}_{N,i})^{1/2}$ and letting SR denote the Sharpe ratio of the portfolio \mathbf{w}_N , we always have

$$\begin{aligned} SR^2 &= \frac{(\mathbf{w}'_N (\boldsymbol{\mu}_N - r_f \mathbf{1}_N))^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \\ &= (SR_1)^2 + (SR_2)^2 - \frac{1}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} (\sigma_2 SR_1 - \sigma_1 SR_2)^2 \\ &\leq (SR_1)^2 + (SR_2)^2. \end{aligned}$$

Finally, strict equality holds if and only if $\sigma_2 SR_1 - \sigma_1 SR_2 = 0$, that is:

$$\frac{\mathbf{w}'_{N,1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}{\mathbf{w}'_{N,1} \mathbf{V}_N \mathbf{w}_{N,1}} = \frac{\mathbf{w}'_{N,2} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}{\mathbf{w}'_{N,2} \mathbf{V}_N \mathbf{w}_{N,2}}.$$

Proof. Defining for simplicity

$$\mu_i - r_f = \mathbf{w}'_{N,i} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \quad \text{and} \quad \sigma_i^2 = \mathbf{w}'_{N,i} \mathbf{V}_N \mathbf{w}_{N,i},$$

we have

$$\begin{aligned} SR^2 &= \frac{(\mu_1 - r_f)^2}{\sigma_1^2} \frac{\sigma_1^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \frac{\sigma_2^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \\ &= \frac{(\mu_1 - r_f)^2}{\sigma_1^2} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \\ &\quad + \left[\frac{(\mu_1 - r_f)^2}{\sigma_1^2} \left(-1 + \frac{\sigma_1^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \right) + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \left(-1 + \frac{\sigma_2^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \right) + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \right]. \end{aligned}$$

Using the orthogonality of $\mathbf{w}_{N,1}$ and $\mathbf{w}_{N,2}$, we have $\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N = \mathbf{w}'_{N,1} \mathbf{V}_N \mathbf{w}_{N,1} + \mathbf{w}'_{N,2} \mathbf{V}_N \mathbf{w}_{N,2} = \sigma_1^2 + \sigma_2^2$, so that the term in square brackets can be rewritten as

$$\begin{aligned} & - \frac{(\mu_1 - r_f)^2}{\sigma_1^2} \frac{\sigma_2^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} - \frac{(\mu_2 - r_f)^2}{\sigma_2^2} \frac{\sigma_1^2}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} + 2 \frac{(\mu_1 - r_f)(\mu_2 - r_f)}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \\ &= \frac{1}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \left(-(\mu_1 - r_f)^2 \frac{\sigma_2^2}{\sigma_1^2} - (\mu_2 - r_f)^2 \frac{\sigma_1^2}{\sigma_2^2} + 2(\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right) \end{aligned}$$

$$= -\frac{1}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \left((\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right)^2.$$

Hence,

$$\begin{aligned} \text{SR}^2 &= \frac{(\mu_1 - r_f)^2}{\sigma_1^2} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} - \frac{1}{\mathbf{w}'_N \mathbf{V}_N \mathbf{w}_N} \left((\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right)^2 \\ &\leq \frac{(\mu_1 - r_f)^2}{\sigma_1^2} + \frac{(\mu_2 - r_f)^2}{\sigma_2^2} = (\text{SR}_1)^2 + (\text{SR}_2)^2. \end{aligned}$$

Equality holds if and only if

$$\left((\mu_1 - r_f) \frac{\sigma_2}{\sigma_1} - (\mu_2 - r_f) \frac{\sigma_1}{\sigma_2} \right)^2 = 0,$$

which, in turn, can be rearranged as

$$\frac{(\mu_1 - r_f)}{\sigma_1^2} = \frac{(\mu_2 - r_f)}{\sigma_2^2}.$$

■

C.2 Extension of Roll (1980)

Roll (1980) shows that, in the absence of a risk-free rate, for any *inefficient* portfolio one can identify the subspace of portfolios that are orthogonal to this portfolio with minimum variance. That is, corresponding to any inefficient portfolio, the number of zero-beta portfolios is infinite—one for each level of target mean. If the portfolio is efficient, then the subspace shrinks to a single point; that is, there is a unique zero-beta portfolio. In order to interpret our findings, we extend the result in Roll (1980) to the case in which investors can invest also in a risk-free asset.

Theorem C2 (Extension of Roll (1980) to the case with a risk-free asset). *Let \mathbf{w}_N^x be any, possibly inefficient, portfolio. Let \mathbf{w}_N^z be the portfolio that satisfies*

$$\min \frac{1}{2} (\mathbf{w}_N^z)' \mathbf{V}_N \mathbf{w}_N^z \quad \text{s.t.} \quad (\mathbf{w}_N^x)' \mathbf{V}_N \mathbf{w}_N^z = 0$$

and

$$\boldsymbol{\mu}'_N \mathbf{w}_N^z + (1 - \mathbf{1}'_N \mathbf{w}_N^z) r_f = \mu^z,$$

for a given target mean μ^z . Then,

$$\mathbf{w}_N^z = \left(\mathbf{w}_N^x, \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right) \begin{pmatrix} (\sigma^x)^2 & \mu^x - r_f \\ \mu^x - r_f & (\text{SR}^{mv})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^z - r_f \end{pmatrix},$$

where $\left(\mathbf{w}_N^x, \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right)$ is the $N \times 2$ matrix obtained by joining the $N \times 1$ vector of portfolio weights \mathbf{w}_N^x with the $N \times 1$ vector $\mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$.

Proof. We adapt Roll's (1980) proof of the main theorem. The Lagrangian for our problem is

$$L(\mathbf{w}_N^z, \lambda_1, \lambda_2) = (\mathbf{w}_N^z)' \mathbf{V}_N \mathbf{w}_N^z - \lambda_1 ((\mathbf{w}_N^x)' \mathbf{V}_N \mathbf{w}_N^z) - \lambda_2 (\boldsymbol{\mu}'_N \mathbf{w}_N^z + (1 - \mathbf{1}'_N \mathbf{w}_N^z) r_f - \mu^z),$$

with first-order conditions

$$2\mathbf{V}_N \mathbf{w}_N^z = \left(\mathbf{V}_N \mathbf{w}_N^x, (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right) \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$

Pre-multiplying both sides by $2^{-1} \left(\mathbf{V}_N \mathbf{w}_N^x, (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right)' \mathbf{V}_N^{-1}$ gives

$$\begin{pmatrix} 0 \\ \mu^z - r_f \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (\sigma^x)^2 & \mu^x - r_f \\ \mu^x - r_f & (SR^{mv})^2 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$

Substituting out for $\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$ concludes the proof. ■

Observe that when \mathbf{w}_N^x is efficient, then $\mathbf{w}_N^z = \mathbf{0}$, which implies that the zero-beta portfolio to \mathbf{w}_N^x is the portfolio that invests 100% in the risk-free asset. In fact substituting $\mathbf{w}_N^x = \gamma^{-1} \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$ into the first constraint gives $0 = (\mathbf{w}_N^x)' \mathbf{V}_N \mathbf{w}_N^z = \gamma^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{w}_N^z = \gamma^{-1} (\mu_z - r_f)$, where the last equality is due to the second constraint. Therefore one obtains $\mu_z = r_f$ which, by no-arbitrage, implies $\mathbf{w}_N^z = \mathbf{0}$.

Recall the well-known result that the entire efficient frontier of risky assets can be generated from holding any two efficient portfolios. However, one can show that the efficient frontier of risky assets can also be generated by holding two *inefficient* portfolios, as long as one is the minimum-variance orthogonal portfolio of the other.

Theorem C3 (Decomposing weights of mean-variance portfolio). *Suppose that the vector of asset returns, \mathbf{r}_t , satisfies Assumptions 1 and 2. Then for any finite $N > K$ and any $\mu^* > r_f$:*

(i) *The mean-variance portfolio weights, with a target mean of μ^* , satisfy the following decomposition:*

$$\mathbf{w}_N^{mv} = \phi^\alpha \mathbf{w}_N^\alpha + \phi^\beta \mathbf{w}_N^\beta,$$

where

$$\mathbf{w}_N^\alpha = \frac{1}{\gamma^\alpha} \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N, \quad \text{and}$$

$$\mathbf{w}_N^\beta = \frac{1}{\gamma^\beta} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}},$$

$$\text{with } \gamma^\alpha = \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{\mu^* - r_f}, \gamma^\beta = \gamma - \gamma^\alpha = \frac{\check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}}{\mu^* - r_f}, \phi^\alpha = \frac{\gamma^\alpha}{\gamma} = \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)},$$

$$\phi^\beta = \frac{\gamma^\beta}{\gamma} = 1 - \phi^\alpha, \text{ with } \gamma = \frac{(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)}{\mu^* - r_f}. \quad 32$$

(ii) Furthermore, the portfolios \mathbf{w}_N^α and \mathbf{w}_N^β satisfy the orthogonality condition,

$$(\mathbf{w}_N^\alpha)' \mathbf{V}_N \mathbf{w}_N^\beta = (\mathbf{w}_N^\alpha)' \boldsymbol{\Sigma}_N \mathbf{w}_N^\beta = 0.$$

Moreover, \mathbf{w}_N^β is the minimum-variance portfolio that is orthogonal to \mathbf{w}_N^α and vice versa.

(iii) Finally, we have two-fund separation: the possibly inefficient portfolios \mathbf{w}_N^α and \mathbf{w}_N^β can generate all the portfolios on the efficient mean-variance frontier of risky assets.

Remark C3.1. One can interpret γ^α as the ratio of the share of the contribution of \mathbf{w}_N^α to the expected return on the mean-variance portfolio with unit risk aversion, $(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$, over the target excess mean return, $\mu^* - r_f$.

Proof: (i) Given that $\boldsymbol{\mu}_N - r_f \mathbf{1}_N = \check{\boldsymbol{\alpha}}_N + \mathbf{B}_N \check{\boldsymbol{\lambda}}$,

$$\begin{aligned} \mathbf{w}_N^{\text{mv}} &= \frac{1}{\gamma} \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \\ &= \frac{1}{\gamma} \mathbf{V}_N^{-1} \check{\boldsymbol{\alpha}}_N + \frac{1}{\gamma} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\ &= \frac{\gamma^\alpha}{\gamma} \frac{1}{\gamma^\alpha} \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N + \frac{\gamma^\beta}{\gamma} \frac{1}{\gamma^\beta} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\ &= \phi^\alpha \mathbf{w}_N^\alpha + \phi^\beta \mathbf{w}_N^\beta. \end{aligned}$$

(ii) Moreover,

$$\mathbf{w}_N^\alpha' \mathbf{V}_N \mathbf{w}_N^\beta = \frac{1}{\gamma^\beta} \mathbf{w}_N^\alpha' \mathbf{V}_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} = \frac{1}{\gamma^\beta} \mathbf{w}_N^\alpha' \mathbf{B}_N \check{\boldsymbol{\lambda}} = \frac{1}{\gamma^\alpha \gamma^\beta} \check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \mathbf{B}_N \check{\boldsymbol{\lambda}} = 0,$$

because $\boldsymbol{\Sigma}_N^+$ is orthogonal to \mathbf{B}_N . Similarly,

$$\begin{aligned} \mathbf{w}_N^\alpha' \boldsymbol{\Sigma}_N \mathbf{w}_N^\beta &= \frac{1}{\gamma^\beta} \mathbf{w}_N^\alpha' \boldsymbol{\Sigma}_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \\ &= \frac{1}{\gamma^\beta} \mathbf{w}_N^\alpha' \mathbf{B}_N (\mathbf{I}_K - (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N) \check{\boldsymbol{\lambda}} \\ &= \frac{1}{\gamma^\alpha \gamma^\beta} \check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \mathbf{B}_N (\mathbf{I}_K - (\boldsymbol{\Omega}^{-1} + \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \boldsymbol{\Sigma}_N^{-1} \mathbf{B}_N) \check{\boldsymbol{\lambda}} \\ &= 0. \end{aligned}$$

³²One can interpret γ^α as the ratio of the share of the contribution of \mathbf{w}_N^α to the expected return on the mean-variance portfolio with unit risk aversion, $(\boldsymbol{\mu}_N - r_f \mathbf{1}_N)' \mathbf{V}_N^{-1} (\boldsymbol{\mu}_N - r_f \mathbf{1}_N)$, over the target excess mean return, $\mu^* - r_f$. The role of the ϕ^α and ϕ^β coefficients is to ensure that the \mathbf{w}_N^α and \mathbf{w}_N^β portfolios achieve the same target mean return as \mathbf{w}_N^{mv} , which is μ^* .

We now show that \mathbf{w}_N^α and \mathbf{w}_N^β are the minimum-variance portfolios orthogonal to one another. This is accomplished by showing that these portfolio weights satisfy the result of Lemma C2. In particular, we need to verify that \mathbf{w}_N^β satisfies

$$\mathbf{w}_N^\beta = \left(\mathbf{w}_N^\alpha, \mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right) \begin{pmatrix} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^\beta - r_f \end{pmatrix}.$$

Simple calculations leads to

$$\begin{aligned} & \begin{pmatrix} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^\beta - r_f \end{pmatrix} \\ &= \frac{1}{\left((\text{SR}^{\text{mv}})^2 \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} - \frac{(\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N)^2}{(\gamma^\alpha)^2} \right)} \begin{pmatrix} (\text{SR}^{\text{mv}})^2 & -\frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{\gamma^\alpha} \\ -\frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{\gamma^\alpha} & \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \frac{\check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}}{\gamma^\beta} \\ &= \frac{1}{\left((\text{SR}^{\text{mv}})^2 \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} - \frac{(\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N)^2}{(\gamma^\alpha)^2} \right)} \begin{pmatrix} -\frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{\gamma^\alpha} \\ \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} \end{pmatrix} \frac{\check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}}{\gamma^\beta}. \end{aligned}$$

Given that

$$\left(\frac{\boldsymbol{\Sigma}_N^{-1} \check{\boldsymbol{\alpha}}_N}{\gamma^\alpha}, \mathbf{V}_N^{-1}(\check{\boldsymbol{\alpha}}_N + \mathbf{B}_N \check{\boldsymbol{\lambda}}) \right) \begin{pmatrix} -\frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{\gamma^\alpha} \\ \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} \end{pmatrix} = \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}$$

and

$$\left((\text{SR}^{\text{mv}})^2 \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} - \frac{(\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N)^2}{(\gamma^\alpha)^2} \right) = \frac{1}{(\gamma^\alpha)^2} (\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N) (\check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}),$$

one finally obtains

$$\begin{aligned} & \left(\mathbf{w}_N^\alpha, \mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right) \begin{pmatrix} (\sigma^\alpha)^2 & \mu^\alpha - r_f \\ \mu^\alpha - r_f & (\text{SR}^{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^\beta - r_f \end{pmatrix} \\ &= \frac{1}{\frac{1}{(\gamma^\alpha)^2} (\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N) (\check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}})} \frac{\check{\boldsymbol{\alpha}}_N' \boldsymbol{\Sigma}_N^+ \check{\boldsymbol{\alpha}}_N}{(\gamma^\alpha)^2} \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}} \frac{\check{\boldsymbol{\lambda}}' \mathbf{B}'_N \mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}}{\gamma^\beta} \\ &= \frac{\mathbf{V}_N^{-1} \mathbf{B}_N \check{\boldsymbol{\lambda}}}{\gamma^\beta} = \mathbf{w}_N^\beta. \end{aligned}$$

Along the same lines, it follows that \mathbf{w}_N^α satisfies

$$\mathbf{w}_N^\alpha = \left(\mathbf{w}_N^\beta, \mathbf{V}_N^{-1}(\boldsymbol{\mu}_N - r_f \mathbf{1}_N) \right) \begin{pmatrix} (\sigma^\beta)^2 & \mu^\beta - r_f \\ \mu^\beta - r_f & (\text{SR}^{\text{mv}})^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu^\alpha - r_f \end{pmatrix}.$$

(iii) Given the result in (i), it suffices to recognize that \mathbf{w}_N^α and \mathbf{w}_N^β are inefficient. This follows as long as both $\check{\boldsymbol{\alpha}}_N \neq 0$ and $\check{\boldsymbol{\lambda}} \neq 0$. In fact, given that one is the minimum-variance orthogonal portfolio to the other, if one was efficient, then by Theorem C2, its minimum-variance orthogonal portfolio exposure to the risky assets must be the zero vector. However, by part (i), their linear combination always spans the efficient frontier. ■

C.3 APT restriction in terms of projection errors

In the lemma below, we show that the APT restriction can be expressed as a function of either the projection errors $\check{\alpha}_N$ or their (element by element) limit α_N .

Theorem C4 (Equivalence of APT constraint in terms of $\check{\alpha}_N$ and α_N).

$$\check{\alpha}_N \Sigma_N^{-1} \check{\alpha}_N \leq \delta < \infty \text{ implies } \alpha_N \Sigma_N^{-1} \alpha_N \leq \delta < \infty.$$

Proof. Note that from (3) and (4) we have

$$\begin{aligned} \check{\alpha}_N &= (\mu_N - r_f \mathbf{1}_N) - \mathbf{B}_N \check{\lambda} \\ &= (\mathbf{I}_N - \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1}) (\alpha_N + \mathbf{B}_N \lambda) \\ &= (\mathbf{I}_N - \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1}) \alpha_N. \end{aligned}$$

Because

$$\Sigma_N^{-1} \check{\alpha}_N = \Sigma_N^{-1} (\mathbf{I}_N - \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1}) \alpha_N = \Sigma_N^+ \alpha_N,$$

it follows that

$$\begin{aligned} \check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N &= \check{\alpha}'_N \Sigma_N^{-1} \Sigma_N \Sigma_N^{-1} \check{\alpha}_N \\ &= \alpha'_N \Sigma_N^+ \Sigma_N \Sigma_N^+ \alpha_N \\ &= \alpha'_N \Sigma_N^+ \alpha_N, \end{aligned} \tag{C1}$$

where Σ_N^+ is defined in (13). Therefore, the condition in (C2) below,

$$\check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N \leq \delta < \infty, \tag{C2}$$

implies from (C1) that

$$\check{\alpha}'_N \Sigma_N^{-1} \check{\alpha}_N = \alpha'_N \Sigma_N^+ \alpha_N = \alpha'_N \Sigma_N^{-1} \alpha_N - \alpha'_N \Sigma_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} \alpha_N \leq \delta < \infty. \tag{C3}$$

■

Observe that because $\alpha'_N \Sigma_N^+ \alpha_N \geq 0$, therefore (C3) implies that

$$0 \leq \alpha'_N \Sigma_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} \alpha_N \leq \alpha'_N \Sigma_N^{-1} \alpha_N \leq \delta < \infty,$$

implying that (5), as well as the equation below, hold by no-arbitrage:

$$\alpha'_N \Sigma_N^{-1} \mathbf{B}_N (\mathbf{B}'_N \Sigma_N^{-1} \mathbf{B}_N)^{-1} \mathbf{B}'_N \Sigma_N^{-1} \alpha_N \leq \delta < \infty.$$

References

- AIT-SAHALIA, Y., AND M. W. BRANDT (2001): “Variable Selection for Portfolio Choice,” *Journal of Finance*, 56(1), 1297–1351.
- AL-NAJJAR, N. I. (1998): “Factor Analysis and Arbitrage Pricing in Large Asset Economies,” *Journal of Economic Theory*, 78(2), 231–262.
- ANDERSON, T. M. (1984): *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York.
- ANDREWS, D. (1999): “Estimation When a Parameter is on a Boundary,” *Econometrica*, 67(6), 1341–1383.
- ANG, A. (2014): *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press.
- BLACK, F. (1995): “Estimating Expected Return,” *Financial Analysts Journal*, pp. 168–171.
- BRANDT, M. W. (1999): “Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach,” *Journal of Finance*, 54, 1609–1646.
- (2004): “Portfolio Choice Problem,” in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia, and L. Hansen. Amsterdam: Elsevier Science.
- BRANDT, M. W., P. SANT-CLARA, AND R. VALKANOV (2009): “Parametric Portfolio Policies: Exploiting Characteristics in the Cross Section of Equity Returns,” *Review of Financial Studies*, p. Forthcoming.
- CERREIA-VIOGLIO, S., F. MACCHERONI, M. MARINACCI, AND L. MONTRUCCHIO (2011): “Uncertainty Averse Preferences,” *Journal of Economic Theory*, 146(4), 1275–1330.
- CHAMBERLAIN, G. (1983): “Funds, Factors and Diversification in Arbitrage Pricing Models,” *Econometrica*, 51, 1305–1324.
- CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): “Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets,” *Econometrica*, 51, 1281–1304.
- CHEN, Z., AND L. G. EPSTEIN (2002): “Ambiguity, Risk, and Asset Returns in Continuous Time,” *Econometrica*, 70(4), 1403–1443.
- CHRISTENSEN, T. M. (2017): “Nonparametric Stochastic Discount Factor Decomposition,” *Econometrica*, 85(5), 1501–1536.
- COCHRANE, J. H. (2005): *Asset Pricing (revised edition)*. Princeton University Press, Princeton, New Jersey.
- (2014): “A Mean-Variance Benchmark for Intertemporal Portfolio Theory,” *Journal of Finance*, 69(1), 1–49.
- CONNOR, G., L. GOLDBERG, AND R. A. KORAJCZYK (2010): *Portfolio Risk Analysis*. Princeton University Press, Princeton, New Jersey.

- DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009): “A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms,” *Management Science*, 55(5).
- DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009): “Optimal Versus Naïve Diversification: How Inefficient is the 1/N Portfolio Strategy?,” *Review of Financial Studies*, 22, 1915–1953.
- DOW, J., AND S. R. D. C. WERLANG (1992): “Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio,” *Econometrica*, 60(1), 197–204.
- EPSTEIN, L. G., AND M. SCHNEIDER (2010): “Ambiguity and Asset Markets,” *Annual Review of Financial Economics*, 2, 315–346.
- EPSTEIN, L. G., AND T. WANG (1994): “Intertemporal Asset Pricing Under Knightian Uncertainty,” *Econometrica*, 62(2), 283–322.
- FROST, P., AND E. SAVARINO (1986): “An Empirical Bayes Approach to Efficient Portfolio Selection,” *Journal of Financial and Quantitative Analysis*, 21, 293–305.
- GAGLIARDINI, P., E. OSSOLA, AND O. SCAILLET (2016): “Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets,” *Econometrica*, 84(3), 985–1046.
- (2017): “A Diagnostic Criterion for Approximate Factor Structure,” Working Paper, University of Lugano.
- GARLAPPI, L., R. UPPAL, AND T. WANG (2007): “Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach,” *Review of Financial Studies*, 20(1), 41–81.
- GILBOA, I., AND M. MARINACCI (2016): “Ambiguity and the Bayesian Paradigm,” in *Readings in Formal Epistemology: Sourcebook*, ed. by H. Arló-Costa, V. F. Hendricks, and J. van Benthem, pp. 385–439. Springer International Publishing, Cham.
- GILBOA, I., AND D. SCHMEIDLER (1989): “Maxmin Expected Utility Theory with Non-Unique Prior,” *Journal of Mathematical Economics*, 18, 141–153.
- GREEN, R., AND B. HOLLIFIELD (1992): “When Will Mean-Variance Efficient Portfolios be Well Diversified?,” *Journal of Finance*, 47, 1785–1809.
- GUIDOLIN, M., AND F. RINALDI (2009): “A Simple Model of Trading and Pricing Risky Assets under Ambiguity: Any Lessons for Policy Makers,” Working Paper, Federal Reserve Board, St. Louis.
- HANSEN, L. P., AND T. J. SARGENT (1999): “Wanting Robustness in Macroeconomics,” Working Paper, University of Chicago.
- HANSEN, L. P., AND T. J. SARGENT (2008): *Robustness*. Princeton University Press, Princeton, NJ.
- HASTIE, T., R. TIBSHIRANI, AND M. WAINWRIGHT (2015): *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.

- HUBERMAN, G. (1982): “A Simple Approach to the Arbitrage Pricing Theory,” *Journal of Economic Theory*, 28, 183–191.
- INGERSOLL, J. (1984): “Some Results in the Theory of Arbitrage Pricing,” *Journal of Finance*, 39, 1021–1039.
- JAGANNATHAN, R., AND T. MA (2003): “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps,” *Journal of Finance*, 58, 1651–1684.
- JOBSON, J. D., AND R. KORKIE (1980): “Estimation for Markowitz Efficient Portfolios,” *Journal of the American Statistical Association*, 75, 544–554.
- JORION, P. (1986): “Bayes-Stein Estimation for Portfolio Analysis,” *The Journal of Financial and Quantitative Analysis*, 21(3), 279–292.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2005): “A Smooth Model of Decision Making Under Ambiguity,” *Econometrica*, 73(6), 1849–1892.
- MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2006): “Ambiguity Aversion, Malevolent Nature, and the Variational Representation of Preferences,” *Econometrica*, 74, 1447–1498.
- MACKINLAY, A. C., AND Ľ. PÁSTOR (2000): “Asset Pricing Models: Implications for Expected Returns and Portfolio Selection,” *Review of Financial Studies*, 13(4), 883–916.
- MERTON, R. C. (1980): “On Estimating the Expected Return on the Market: An Exploratory Investigation,” *Journal of Financial Economics*, 8, 323–361.
- PÁSTOR, Ľ. (2000): “Portfolio Selection and Asset Pricing Models,” *Journal of Finance*, 55, 179–223.
- PÁSTOR, Ľ., AND R. F. STAMBAUGH (2000): “Comparing Asset Pricing Models: An Investment Perspective,” *Journal of Financial Economics*, pp. 335–381.
- PETTENUZZO, D., A. G. TIMMERMANN, AND R. VALKANOV (2014): “Forecasting Stock Returns under Economic Constraints,” *Journal of Financial Economics*, 114(3), 517–553.
- RENAULT, E., T. VAN DER HEIJDEN, AND B. WERKER (2017): “Arbitrage Pricing Theory for Idiosyncratic Variance Factors,” Available at SSRN: <https://ssrn.com/abstract=3065854>.
- ROLL, R. (1977): “A Critique of the Asset Pricing Theory’s Tests Part I: On Past and Potential Testability of the Theory,” *Journal of Financial Economics*, 4(2), 129–176.
- (1980): “Orthogonal Portfolios,” *Journal of Financial and Quantitative Analysis*, 15(5), 1005–1023.
- ROSS, S. (1976): “The Arbitrage Theory of Capital Asset Pricing,” *Journal of Economic Theory*, 13, 341–360.
- ROSS, S. A. (1977): “Return, Risk, and Arbitrage,” in *Risk and Return in Finance*, ed. by I. Friend, and J. Bicksler. Ballinger, Cambridge, MA.

- SAVAGE, L. J. (1954): *The Foundations of Statistics*. John Wiley & Sons.
- TREYNOR, J. L., AND F. BLACK (1973): “How to Use Security Analysis to Improve Portfolio Selection,” *The Journal of Business*, 46(1), 66.
- TROJANI, F., AND P. VANINI (2002): “A Note on Robustness in Merton’s Model of Intertemporal Consumption and Portfolio Choice,” *Journal of Economic Dynamics and Control*, 26, 423–435.
- UPPAL, R., AND T. WANG (2003): “Model Misspecification and Underdiversification,” *Journal of Finance*, 58(6), 2465–2486.